Cognition 120 (2011) 341–349

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Where science starts: Spontaneous experiments in preschoolers' exploratory play

Claire Cook [a], Noah D. Goodman [b], Laura E. Schulz [a,*]

[a] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States
[b] Department of Psychology, Stanford University, United States

ARTICLE INFO

Article history:
Received 9 April 2010
Revised 26 February 2011
Accepted 15 March 2011
Available online 10 May 2011

Keywords:
Probabilistic models
Expected information gain
Theory theory
Preschoolers
Causal learning
Exploratory play
Interventions
Grounded causal models

ABSTRACT

Probabilistic models of expected information gain require integrating prior knowledge about causal hypotheses with knowledge about possible actions that might generate data relevant to those hypotheses. Here we looked at whether preschoolers (mean: 54 months) recognize "action possibilities" (affordances) in the environment that allow them to isolate variables when there is information to be gained. By manipulating the physical properties of the stimuli, we were able to affect the degree to which candidate variables could be isolated; by manipulating the base rate of candidate causes, we were able to affect the potential for information gain. Children's exploratory play was sensitive to both manipulations: given unambiguous evidence children played indiscriminately and rarely tried to isolate candidate causes; given ambiguous evidence, children both selected (Experiment 1) and designed (Experiment 2) informative interventions.

1. Introduction

Cooks Illustrated recently published an article (*Cracking the code to chewy brownies*; March, 2010) on how to make homemade brownies with the texture and "shiny, crisp, crackly top" of boxed mixes. A food editor suggested that the testers consider the proportion of solid (saturated) to liquid (unsaturated) fat. The testers proceeded to hold the total fat constant while varying the ratio the fats until tasters judged that they had achieved the perfect brownie.

The article illustrates not only what you need to know to bake a brownie, but also much of what you need to know to conduct an informative experiment. You need to know when there is information to be gained (e.g., that there is some uncertainty about what makes brownies taste good); you need a theory about which variables are relevant (that there are different kinds of fats with different properties); you need to know how to distinguish the causal role of these variables (by changing one factor at a time), and you need to know what features of the environment can be manipulated to let you test these factors (that fats can be mixed in different combinations).

Experimentation thus calls on a diverse range of skills. Arguably however, experimentation is an arcane practice, of interest primarily to scientists and America's Test Kitchen. There are at least two reasons however, to think that some principles of experimental design are central to everyday cognition.

First, daily life frequently presents us with competing causal hypotheses. If we cannot unlock a door, it makes sense to rotate the key or try a new key but not to try a new key in a new position. Even lay adults seem to recognize the value of interventions that isolate variables.

Second, contemporary theories of cognitive development have been profoundly shaped by an analogy to science. The *theory theory*, the idea that everyday knowledge has the structural, functional, and dynamic properties of scientific

* Corresponding author. Address: Brain & Cognitive Sciences Department, 77 Massachusetts Ave., 46-4011b, Massachusetts Institute of Technology, Cambridge, MA 02138, 617-253-7957, United States.

E-mail address: lschulz@mit.edu (L.E. Schulz).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.03.003

theories, has been among the most influential accounts of cognitive development in recent decades (Carey, 1987, Carey, 2000; Gelman, 2003; Gopnik & Meltzoff, 1997; Gopnik & Schulz, 2007; Keil, 1989; Wellman, 1990). Research in this tradition attests to the structural, dynamic, and functional similarities between children's knowledge and scientific theories, pointing out that both kinds of knowledge are abstract, coherent, and causal; both involve an interaction between evidence and prior beliefs, and both support prediction, explanation, and intervention (see Gopnik and Meltzoff (1997) for exposition and review).

The "child as scientist" account would seem to predict that an additional functional feature of theories – the ability to support informative exploration – should also emerge in early childhood. However, evidence for this seemingly fundamental point of comparison between science and cognitive development, the dynamic by which new knowledge is acquired, has been strikingly mixed. Indeed, education research looking at the relationship between self-guided exploration and science learning has found evidence *against* the claim that children "learn by doing." Studies suggest that students have a poor metacognitive understanding of principles of experimental design, difficulty designing controlled interventions, and difficulty anticipating the type of evidence that would support or undermine causal hypotheses (Inhelder & Piaget, 1958; Klahr & Nigam, 2004; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Koslowski, 1996; Masnick & Klahr, 2003).

Research in science education however, typically investigates students' understanding of real world phenomena (e.g., density, balance relations, etc.). In such contexts, children's reliance on domain-specific prior beliefs may mask their formal reasoning abilities (Koslowski, 1996; Kuhn, 1989; Kushnir & Gopnik, 2005; Schulz, Bonawitz, & Griffiths, 2007; Schulz & Gopnik, 2004; Sobel & Munro, 2009). Additionally, students are often tested on relatively complex, multivariate problems (e.g., Kuhn, 1989; Masnick & Klahr, 2003). Such problems are appropriate for investigating factors that could affect classroom performance but may underestimate children's causal reasoning in simpler contexts.

Developmental studies provide stronger grounds for optimism about children's ability to design informative interventions. Work in fields ranging from perception to motor learning to industrial design (e.g., Adolph, Eppler, & Gibson, 1993; Berger, Adolph, & Lobo, 2005; Brown, 1990; Lockman, 2000; Norman, 1988, 1999) suggests that learners discover action possibilities or *affordances* (Gibson, 1977) in the environment through exploration. Research suggests for instance that toddlers inspect the length and ends of rakes when they need a tool to reach a distant object (Brown, 1990), and the rigidity of handrails when they need to cross narrow bridges (Berger et al., 2005). Similarly, when access to a toy or food is obstructed, toddlers, non-human primates, and even corvids can perform novel interventions to gain information and achieve their goals (Brauer, Kaminski, Reidel, Call, & Tomasello, 2006; Emery & Clayton, 2004; Hood, Carey, & Prasada, 2000; Mendes, Hanus, & Call, 2007; Stulp, Emery, Verhulst, & Clayton, 2009). However, children can learn object functions without designing experiments; the ability to intervene on physical features of the environment to gain information does not necessarily entail the ability to intervene when information is unknown because of formal properties of the evidence (e.g., because causal variables are confounded).

The strongest evidence that children may understand some formal principles underlying experimental design comes from research looking at children's causal reasoning. Studies suggest, for instance, that preschoolers understand patterns of co-variation well enough to distinguish genuine causes from spurious associations: if two variables together generate an effect but only one variable generates the effect independently, children conclude that the other variable is not a cause (Gopnik, Sobel, Schulz, & Glymour, 2001; Kushnir & Gopnik, 2005, 2007; Schulz & Gopnik, 2004). Children's causal judgments are also sensitive to the base rate of candidate causes. When the status of a causal variable is ambiguous, preschoolers are more likely to believe it is causal when causes are common than when they are rare (Sobel, Tenenbaum, & Gopnik, 2004). Moreover, preschoolers can draw accurate inferences not only from observed evidence but also from evidence they generate (by chance) in exploratory play (Schulz, Gopnik, & Glymour, 2007). Finally, two recent studies (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007) suggest that children's exploratory play is affected by the ambiguity of the evidence they observe; given confounded or un-confounded evidence about which of two variables controls which of two effects, preschoolers' selectively explore confounded evidence. Critically however, selective exploration of confounded evidence is advantageous even if children explore randomly (with no understanding of how to isolate variables): the more different actions children perform, the better their odds of generating informative data.

Thus despite evidence for the sophistication of children's causal reasoning, previous work falls short of suggesting that children design interventions that respect principles of experimental design: children might both learn from informative evidence (Gopnik et al., 2001; Schulz & Gopnik, 2004; Schulz et al., 2007; Sobel et al., 2004), and selectively explore uninformative evidence (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007), without any understanding of what it is *about* evidence that makes it informative or uninformative. Specifically, preschoolers might not understand that isolating variables generates informative evidence, or that confounding generates uninformative evidence. Do children understand the kind of interventions that support information gain? Do they selectively perform interventions that isolate competing candidate causes?

Before investigating empirically whether there are contexts in which children can engage in effective experimentation, we might want to know, normatively, when and how they should do so. That is, how should one choose an action, if one's goal is to learn how a causal system works? In answering this question, we are motivated by work on probabilistic models in adult cognitive science, looking at exploration/exploitation trade-offs and decision-making under uncertainty (e.g., Daw, Niv, & Dayan, 2005; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006;

Kaelbling, 1993; Kang et al., 2009; McClure, Daw, & Read Montague, 2003; Sutton & Barto, 1998). Although many of these studies have focused not on richly structured causal domains, but on arbitrary reinforcement learning problems (e.g., multi-armed bandit tasks, stacked card decks, maze-running tasks with hidden rewards, etc.), computational models of the expected information gain associated with different actions establish an important starting point for the current work.

It is optimal (in a decision-theoretic sense) to maximize the expected information gain of an action (see Oaksford & Chater, 1994); in particular, it is optimal to choose the action that will most decrease the uncertainty over the causal structure that relates the variables of interest. Formally, the information gain from observing data *D* after taking action *A* can be represented by:

$$I_g(D,A) = I(P(H|D,A)) - I(P(H)), \quad (1)$$

where $I(p)$ represents the Shannon–Wiener information (Shannon and Weaver, 1948; Wiener, 1948) of distribution p: $I(p) = -\Sigma_i p_i \log_2(p_i)$. The expected information gain from taking action *A* is then:

$$EI_g(A) = \Sigma_i P(D_i|A) I_g(D_i,A), \quad (2)$$

where the probability of an observation given an action is:

$$P(D_i|A) = \Sigma_k P(D_i|A,H_k)P(H_k). \quad (3)$$

In choosing the action which maximizes the expected information gain there are thus two main contributing factors: $P(H)$ and $P(D|A,H)$. These describe where in the system there is information to be learned and how that information can be gathered. The prior distribution $P(H)$ will capture effects of prior knowledge (such as baserates) and earlier evidence (such as confounded observations). As noted, previous research (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007) has indeed shown that children explore a causal system more when there is information to be gained.

The second term influencing optimal information gain describes how actions give rise to observed data, given the causal structure hypothesis. Typically this is assumed to be a simple, direct function: for each object there is an action (e.g. putting it on the detector) that will lead to data that reflects the causal status of that object (e.g. the detector activates if the object is a 'blicket'). Under this assumption, there is no interesting contribution from the $P(D|A,H)$ term; actions will be taken on the objects about which there is the greatest uncertainty.

However, real world scientific situations, and many situations facing children, have a much more complex relationship between available actions and the causal variables of interest. Designing an experiment is often harder than merely choosing what one should try to learn. For instance, if there is not a one-to-one relationship between actions and objects then $P(D|A,H)$ can reflect an inability to isolate some causal variables. In our first experiment, we set up a situation in which there are four objects of uncertain causal status, but only three actions: two which will lead to evidence about two objects individually, and one which leads to (confounded) evidence about a pair of objects. Thus while the contribution of $P(H)$ to EI_g supports exploration of all four objects equally, the incorporation of $P(D|A,H)$ leads to preferential exploration with the separable objects. We test this prediction in Experiment 1.

Perhaps more than any other claim about causal reasoning in young children, the idea that children might be capable of selectively performing informative interventions puts the analogy between children and scientists to the test. This is the claim we investigate here. We hypothesize that when the probability of information gain is high, preschoolers will exploit available affordances to isolate and test causal variables consistent with their folk theories about plausible mechanisms.

By suggesting that preschoolers selectively perform informative interventions, we mean both something *more* than the idea that children learn through exploratory play and something *less* than the idea that children explicitly understand and apply principles of experimental design. We believe children's spontaneous experimentation can be distinguished from trial and error learning or "mere" exploratory behavior by its selectivity. That is, we predict that children will be more likely to perform actions that isolate relevant variables when the probability of information gain is high than when it is low, and that children will be specifically likely to perform actions that isolate relevant variables (rather than simply acting more in general). Children's spontaneous experimentation can be distinguished from a meta-cognitive understanding of experimental design both by its noisy implementation and its fragility. We do not predict that children will perform only informative experiments, that they will perform informative experiments in preference to other playful actions, or that they will perform informative experiments methodically (without redundancy or interruption). Additionally, we believe that children's ability to generate informative actions can be easily compromised by other task demands (e.g., by increasing the number of variables involved or changing the status of those variables with respect to the children's prior beliefs). We presume that bringing the ability to generate informative interventions to bear on tasks of arbitrary complexity requires formal science education.

Here we look at children's ability to design interventions in a simple toy world. We give children base rate information about candidate causes, showing them either that 4 of 4 beads (the *All Beads* condition) or 2 of 4 beads (the *Some Beads* conditions) activate a toy when the beads are placed, one at a time, on top of the toy. We then show both groups of children two pairs of beads. All children learn that one of the bead pairs can be pulled apart into two individual beads, while the other pair is glued together. Finally, children learn that both bead pairs activate the toy (see Fig. 1).

Although in principle only one bead in each pair might be causally effective, the evidence about the bead pair should be relatively unambiguous for children in the *All Beads* condition; the base rate information strongly supports the hypothesis that both beads in both pairs activate the toy. By contrast, the evidence about the bead pairs is genuinely ambiguous for children in the *Some Beads* conditions; the evidence fails to distinguish which bead works (or whether both do). Put another way, when *All Beads*

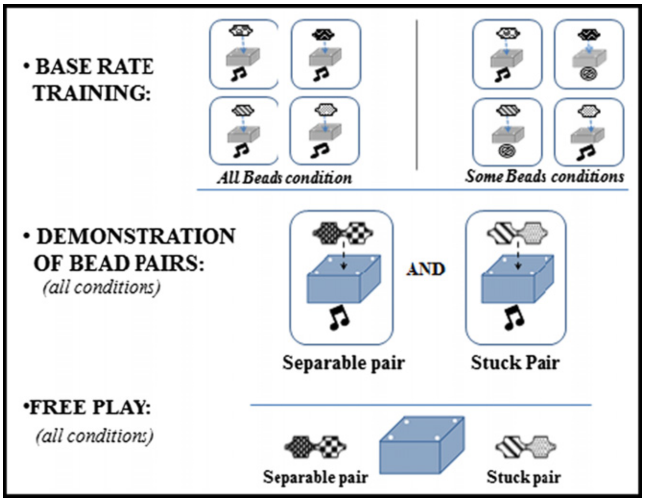

Fig. 1. Procedure used in Experiment 1.

are presumed to be effective, there is relatively little information to be gained about the new beads; when only *Some Beads* are effective (but you do not know which ones are), there is the potential for information gain.

Since evidence suggests that preschoolers expect physical causal relations to respect contact causality (e.g., Kushnir & Gopnik, 2007), we chose that as the folk theory relevant to experiments the children might perform. If children only bring their knowledge of physical causality to bear, they could always place both bead pairs on the toy. However, if children understand that causal variables need to be tested independently, they need to figure out how this can be instantiated with the materials at hand. In particular, they need to look for beads that allow isolated contact with the toy. Only the separable bead pair has the necessary affordances. Moreover, if children specifically separate the beads and place them individually on the toy because of the potential for information gain (rather than because they simply enjoy separating the beads) then children should be more likely to perform this intervention in the *Some Beads* condition than the *All Beads* condition. Finally, if faced with ambiguous evidence children engage in spontaneous experimentation rather than simply extensive free play, the conditions should differ with respect to children's tendency to perform the informative experiment, but not otherwise.

2. Experiment 1

2.1. Method

2.1.1. Participants

Sixty preschoolers (52% girls) were randomly assigned to one of three conditions: an *All Beads* (mean: 53 months; range: 46–63 months), a *Some Beads B* (mean: 55 months; range: 48–61 months) or a *Some Beads AB* (mean: 54 months; range: 46–64 months) condition. The condition names indicate the beads that activated the machine during the Base Rate Training. In the *All Beads* condition, all beads were effective; in both *Some Beads* conditions, half the beads were effective and half were not. During Free Play, children in the *Some Beads* condition could discover either that only the first bead, only the second bead, or, both beads activated the machine. We gave children evidence that only the second bead worked (*Some Beads B* condition) and that both beads worked (*Some Beads AB* condition).[1] Seven participants were excluded and replaced, six due to experimental error (1 in the *All Beads*, 1 in the *Some Beads B* and 4 in the *Some Beads AB* condition) and one (in the *Some Beads B* condition) due to a language barrier.

2.1.2. Materials

Eight 15 cm × 10 cm × 10 cm Fisher–Price connectable beads were used. The beads were shaped and painted so that no two were perceptually identical. Two of the beads were glued together and could not be snapped apart. A 23 cm × 23 cm × 6 cm custom-built machine was also used. The machine had four green LED lights at the top corners. A hidden remote switch, controlled by the experimenter, turned the machine on and off. The top of the machine had a pressure-sensitive platform (21 cm diameter). When the switch was in the on position, the machine played music and lit up when the beads were placed on the platform; the machine turned off as soon as the beads were removed. When the switch was in the off position, the machine was always inert.

2.1.3. Procedure

2.1.3.1. Warm up. The warm-up period was designed to familiarize the child with the stimuli and to reduce any novelty associated with merely snapping the beads to-

[1] Of course, once a child discovers that the first bead she tries does not work (i.e., in the *Some Beads B* condition) she can infer that the second bead does, but we assumed that children would be eager to confirm this. We did not give children evidence that only the first bead worked because (insofar as only the evidence of the first bead could affect children's subsequent actions) it was redundant with the *AB* condition.

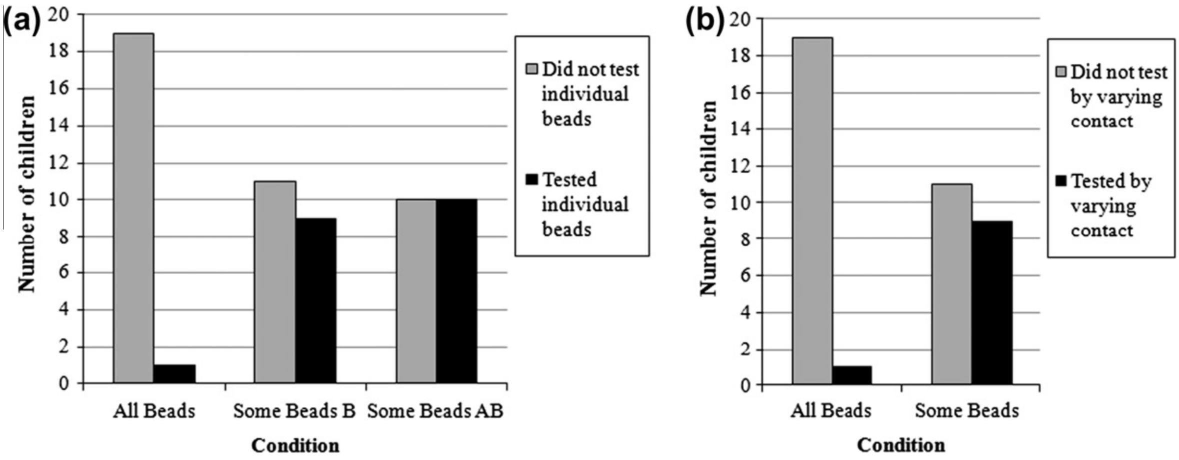

Fig. 2. Number of children in each condition who performed the informative intervention in (a) Experiment 1 and (b) Experiment 2.

gether and apart. The experimenter showed the child four beads and pointed out that each bead looked different. The experimenter showed the child that the beads could be snapped together and pulled apart, and encouraged the child to try. Children were allowed to play with the beads as long as they liked; there was no difference in children's mean play time between conditions (*All Beads*, 32.81 s; *Some Beads B*, 31.91 s; *Some Beads AB*, 28.83 s; $F(2, 25) = 1.38$, $p = ns$) or the mean number of times children snapped apart the beads during the warm-up period (*All Beads*, 6.45; *Some Beads B*, 5.73, *Some Beads AB*, 5.00; $F(2, 25) = .24$, $p = ns$).[2]

2.1.3.2. Base rate training. The experimenter introduced her "special machine" and told the child, "Some things make the machine go, and some things don't make the machine go" (allowing the interpretation in the *All Beads* condition that things other than beads might fail to activate the machine). The experimenter placed each bead on the machine one at a time. In the *All Beads* condition, all the beads activated the machine. In both *Some Beads* conditions, two of the four beads activated the machine and the other two beads did not (either the first and third or first and fourth bead activated, counterbalanced across children).

2.1.3.3. Demonstration of bead pairs and free play. The experimenter removed the training beads and introduced the paired beads. One pair was placed to the right of the machine and the other to the left (counterbalanced). Starting with the pair on the child's right, the experimenter showed the child that one pair was stuck together while the other pair could be pulled apart. The child was allowed to try each pair to verify its status (stuck or separable). The experimenter then placed the pair on the child's right on the machine, and the machine activated. The experimenter took that pair off and placed the other pair on the machine; again the machine activated. Each time the machine activated, the experimenter said, "Wow, look at that. I wonder what makes the machine go?" The experimenter then said, "Go ahead and play" and walked out of the child's line of sight; the child was left to play with the two bead pairs and the machine for 60 s.

2.2. Results and discussion

The critical question was whether children in the *Some Beads* conditions would select the informative action that disambiguates the evidence: separating the separable pair and placing each bead individually on the machine. Of course, children might test each bead individually simply for fun. Thus the dependent measure of interest was whether children performed the specific informative "experiment" (separating the bead pair and testing each member individually on the machine) more often in the *Some Beads* conditions than the *All Beads* condition. Our criteria for the informative experiment were conservative: children had to test each of the two beads separately (i.e., children who placed only one of the beads on the toy were not counted as succeeding).

We first looked at how many children in each condition performed the informative experiment. Results were coded by the first author and a coder blind to both the hypotheses and conditions; intercoder agreement was 95% ($\kappa = .90$). Data from the blind coder was used for all analyses. Reported values are two-tailed tests throughout.

Only one child (5%) in the *All Beads* condition performed the informative experiment. Indeed, the majority of children in the *All Beads* condition (65%) never even pulled the separable bead pair apart. By contrast, 9 children (45%) in the *Some Beads B* and 10 children (50%) in the *Some Beads AB* condition performed the informative experiment. See Fig. 2a. Relative to the *All Beads* condition, children were more likely to perform the informative experiment in both the *Some Beads B* (χ^2 (1, $N = 40$) = 8.53, $p < .005$) and *Some Beads AB* (χ^2 (1, $N = 40$) = 10.16, $p < .005$) condition; performance

[2] Some parents continued to read the consent form during the warm-up period. We only videotaped the warm-up period if the consent form was submitted so this data was available for 17 of 20 children the All Beads condition, 16 of 20 children in the Some Beads B condition and 12 of 20 children in the Some Beads AB condition.

did not differ between the *Some Beads* conditions (χ^2 (1, $N = 40$) = 0.10, $p = ns$).

Children might be more likely to explore ambiguous than unambiguous evidence (consistent with Gweon & Schulz, 2008; Schulz & Bonawitz, 2007) and yet explore broadly, without any insight into potentially informative actions. If so, children might happen upon the informative experiment by chance without specifically choosing the intervention because it is informative. To see whether children differed only with respect to their tendency to perform the informative experiment or whether children simply played more variably overall in the *Some Beads* than the *All Beads* conditions, we coded children's play in four second intervals and categorized each interval into one of seven mutually exclusive categories: playing with the separable beads (snapped apart) on the machine, playing with the separable bead pair (snapped together) on the machine, playing with the stuck bead pair on the machine, playing with three or four beads on the machine, playing with the machine alone, playing with the beads alone, or other (e.g., disengaged). Note that playing with the separable beads snapped on the machine is a component of the informative experiment; however because our criteria were conservative (children had to test each bead in the separable pair individually) the informative experiment does not reduce to any of these categories. All free play actions were coded by two coders, one blind to conditions and the other blind to both the hypotheses and conditions; intercoder agreement was 93% ($K = .85$). Data from the hypothesis-blind coder was used for all analyses.

If children simply explore more broadly given ambiguous evidence, we should expect children in the *Some Beads* conditions to perform actions in more different categories than children in the *All Beads* condition. Although children's play was variable within conditions (i.e., with some children performing actions in few categories and others in many), play across conditions did not differ: children performed actions across a mean of 3.83 different categories in the *All Beads* condition, 4.06 different categories in the *Some Beads B* condition; and a mean of 3.95 different categories in the *Some Beads AB* condition ($F(2, 60) = 1.08$, $p = ns$). We also looked at whether the *Some Beads* and *All Beads* conditions differed in how many children performed any actions other than the target experiment. The only comparison to reach significance was the action that was critical to the target experiment: more children played with the separable beads snapped apart on the toy in the *Some Beads* conditions than in the *All Beads* conditions (63% vs. 18%; χ^2 (1, $N = 40$) = 9.37, $p < .005$). None of the other comparisons reached significance (Fisher's Exact: $p = ns$ throughout).[3]

These results are consistent with the hypothesis that preschool children are sensitive not only to uncertainty, but to the impact of possible actions on information gain (Eqs. (1)–(3)): they spontaneously isolate relevant variables when uncertainty is high (but not when it is low) and specifically perform actions with higher expected information gain. That is, preschoolers not only distinguish relatively ambiguous and unambiguous evidence, they distinguish (and select) potentially informative rather than uninformative actions.

3. Experiment 2

The results of Experiment 1 suggest that preschoolers selectively perform actions that (approximately) maximize expected information gain, integrating information about uncertainty with formal principles underlying experimental design and a sensitivity to relevant affordances. However, in Experiment 1 the target intervention consisted of actions that were already familiar to the children. (They had seen the experimenter place the beads individually on the toy and they had practice snapping the beads apart.) If children could invent novel interventions, rather than merely choose from familiar ones, this would provide stronger evidence of children's ability to act selectively for the purpose of information gain.

In the earlier formal analysis it was assumed that there was a small set of possible actions and that the relationship between actions and outcomes (the term $P(D|A, H)$) was known. In real scientific settings there are a very large number of possible experiments, and the relationship between those experiments, the hypotheses, and the resulting data is not completely known. Knowledge of likely affordances can ameliorate these difficulties by providing prior knowledge about which of many actions is likely to result in evidence about given causal variables.

The behavior of some participants provided an unexpected clue that children might be capable of just such innovation. In Experiment 1, the experimenter always placed the bead pairs horizontally on the toy so that both beads made contact with the platform (see Fig. 1). However, in pilot work, some children performed an action with the stuck pair that they had never observed: they oriented the pair vertically so that only one end of the bead pair touched the toy at a time. The behavior of these children is consistent with the possibility that they were attempting to isolate the causal variables by varying the beads' contact with the toy. In order to look systematically at whether children could invent novel means of disambiguating evidence, we followed up our chance observation. In Experiment 2, we replicated the procedure of Experiment 1 using only the stuck bead pair.

3.1. Method

3.1.1. Participants

Forty preschoolers (45% girls) were randomly assigned to an *All Beads* (mean: 54 months, range: 48–63 months) or a *Some Beads* (mean: 53 months, range: 43–61 months) condition. Because the two *Some Beads* conditions in Experiment 1 did not differ, we counterbalanced the evidence children received during the *Free Play* period, such that for half the children, the first bead they tried failed

[3] Note that because we predicted that the free play would *not* differ except with respect to the target experiment, it was more conservative to test each action as if it were the unique planned comparison than to correct for multiple comparisons. Thus in both Experiments 1 and 2 we use the same criteria for the category intervals as for the target experiment and retain an uncorrected α of .05.

to activate the machine (as in the *Some Beads B* condition of Experiment 1) and for half, both beads activated the machine (as in the *Some Beads AB* condition of Experiment 1). Five children (3 in the *All Beads* condition and 2 in the *Some Beads* condition) were replaced: two due to experimental error, one due to parental interference, and two who did not wish to complete the task.

3.1.2. Materials

The same materials used in Experiment 1 were used in Experiment 2 except that the separable pair of beads was not used.

3.1.3. Procedure

The procedure was identical to the procedure in Experiment 1 except that in the *Demonstration* phase, only the stuck pair of beads was introduced. After the children manipulated the bead pair to determine that it really was stuck together, the experimenter placed the pair on the machine (oriented horizontally, so that both beads made contact, as in Fig. 1). The machine activated. The experimenter said, "Oh, wow, look at that. I wonder what makes the machine go?" The experimenter removed the pair from the machine. The experimenter told the child to "Go ahead and play" and left the child's line of sight; the child was allowed to play for 60 s.

3.2. Results and discussion

The critical question was whether children in the *Some Beads* condition could invent an action consistent with their folk theories about contact causality that could disambiguate the evidence: orienting the beads vertically in order to "isolate" the relevant variables (see Fig. 3). Of course, children might try each end of the bead pair for reasons other than an interest in disambiguating the evidence (i.e., simply for fun). Thus the dependent measure of interest was whether children performed the specific informative experiment more often in the *Some Beads* condition than in the *All Beads* condition. Again, our criteria for the informative experiment were conservative: children had to try each end of the bead pair separately. (Children who touched only a single end were not counted as succeeding.)

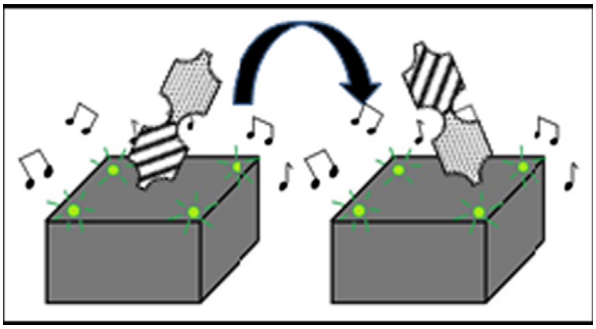

Fig. 3. The informative intervention of interest in Experiment 2, varying the stuck pair's contact with the machine (*shown here: "AB" type evidence; the machine activates for both the first and second tip of the bead pair*).

Results were coded by the first author and a coder blind to both the hypotheses and conditions; intercoder agreement was 96%[4]. Data from the hypothesis-blind coder was used throughout. Only 1 child (5%) in the *All Beads* condition performed the informative experiment. By contrast, 9 children (45%) in the *Some Beads* condition performed the informative experiment (3 children who saw only the second bead activate during free play; 6 children who saw both beads activate). Children were more likely to perform the informative experiment in the *Some Beads* than the *All Beads* condition (χ^2 (1, $N = 40$) = 8.53, $p < .005$) see Fig. 2b. As in Experiment 1, we wanted to look at whether children's play differed broadly between the two conditions or whether their play differed only with respect to the intervention of interest. We coded children's behavior during the free play period at fifteen 4 s intervals into one of five mutually exclusive categories: playing with the entire stuck pair contacting the machine, playing with only one tip of the stuck pair contacting the machine, playing with the beads alone, playing with the machine alone, or other. Again, because our criteria were conservative (children had to test each end of the stuck pair) the informative experiment does not reduce to any of these categories. Results were coded by two coders, one blind to conditions and one blind to both the hypotheses and conditions; intercoder agreement was 86% ($K = .71$). The results of the hypothesis-blind coder were used for all analyses.

If children simply explore more broadly given ambiguous evidence, we should expect children in the *Some Beads* condition to perform actions in more different categories than children in the *All Beads* condition. Again, although children's play was variable within conditions, it did not differ across conditions: children performed actions across a mean of 3.00 different categories in the *Some Beads* condition and 2.78 different categories in the *All Beads* condition; there was no significant difference between conditions, $t(40) = 0.89$, $p = ns$. We also looked at whether there were any differences between the *Some Beads* and *All Beads* conditions in how many children performed any of the other actions, not hypothesized to be of interest. None of the comparisons reached significance (Fisher's Exact one-tailed: $p = ns$ throughout).

In Experiment 2, the relevant intervention was never shown to the child. (Indeed the intervention the children discovered was sufficiently non-obvious that it had not initially occurred to the authors of this paper.) Nonetheless, replicating the results of Experiment 1, children who believed that some beads activated the toy spontaneously attempted to isolate the variables. Subscribing to a principle of contact causality, they varied the beads' contact with the toy so that only a single bead made contact at a time. This behavior did not otherwise occur; children who believed that all the beads activated the toy almost never performed the target action.

The children were of course wrong about the mechanism activating the toy: anything that depressed the platform activated the toy when the switch was "on." Here however, we were interested, not in what children under-

[4] One child's videotape could not be recoded due to technical difficulties.

stood about the actual mechanics of the toy (evidence suggests that even adults have a shallow understanding of most causal mechanisms; Keil, 2003; Rozenblit & Keil, 2002), but in their ability to design potentially informative interventions, consistent with their folk theories about contact causality. The results of Experiment 2 suggest that children not only have this ability but also can exercise it with considerable ingenuity.

4. General discussion

These results suggest that preschoolers distinguish, not only ambiguous and unambiguous evidence but also potentially informative and uninformative interventions. In cases where there was information to be gained, preschoolers spontaneously selected (Experiment 1) and designed (Experiment 2) actions to effectively isolate the relevant variables. Critically, the target experiments were not otherwise part of children's exploratory repertoire; children almost never performed them given unambiguous evidence.

Maximizing the expected gain in information formalizes the "correct" experiment to learn about a given causal system. As our analysis showed, many factors affect the optimal actions: prior knowledge and recent experience enter through the term $P(H)$, while knowledge about possible actions and likely affordances enters through the term $P(D|A,H)$. Importantly, optimal action selection requires integrating these factors into the computation of expected information gain, just as science requires knowing where there is something to be learned and also how to learn it. Our results suggest that children are sensitive to all of these factors and integrate them to guide exploratory play. We believe that these results tighten the analogy to science that has motivated contemporary theories of cognitive development.

This study does not establish however, whether children understood the importance of isolating variables initially or whether they inferred its importance in the course of the experiment. In principle, children might have recognized that they did not know the causal status of the beads during the bead pair demonstration, did know their status during the base rate training and, by comparing (in memory) the factors that differed between the two states, inferred that isolating the beads could support learning. In Experiment 2, the inference is complicated by the fact that children could not simply re-create the state of the beads that obtained when they knew their causal status; children had to make a general inference about the importance of isolating variables and the ways of doing this. Given that the training and bead pair presentation differed in many respects (e.g., different color beads were presented, the experimenter said different things, etc.), it may be attributing more to children to credit them with the ability to infer the relevant distinction without prior knowledge than to credit them with an initial understanding of the importance of isolating variables. However, on either account, these results suggest that preschoolers attend to the kinds of evidence that distinguish states of knowledge from states of uncertainty, and generate novel interventions that isolate variables and maximize the potential for information gain.

This is not to suggest that all aspects of experimental design can be reduced to everyday exploration. As suggested in the Introduction, children might understand the utility of isolating variables in simple contexts without any meta-cognitive understanding that isolating candidate causes *in general* disambiguates confounded data. The ability to bring common principles of experimental design to bear on any task, regardless of the number of variables involved and the status of those variables with respect to their prior beliefs, requires an explicit awareness of the principles of experimental design that is, we presume, the exclusive purview of formal science. Even in everyday life, the actions that support information gain are presumably less obvious than in our study where the presentation of variables was carefully controlled. Further research might investigate whether children can generate informative interventions in more ecologically valid contexts.

Much as science goes beyond simple experiments, so too does exploratory play. Exploratory play is a complex phenomenon, presumably subserving a range of functions other than generating informative evidence (sensorimotor stimulation, the acquisition of new skills, etc.), and affected by a range of factors other than the ambiguity of evidence (the child's interests, temperament, energy level, upbringing, etc.). However, to the extent that children acquire causal knowledge through exploration, the current results begin to bridge the gap between scientific inquiry and child's play.

Acknowledgments

This research was supported by an NSF Faculty Early Career Development Award, a John Templeton Foundation Award and a James S. McDonnell Foundation Collaborative Interdisciplinary Grant on Causal Reasoning to LS. Thanks to Cheryl Kwinn, Sarah Gumlak, Ali Horowitz, and Caroline Huang for their help in data collection and coding and to members of the Early Childhood Cognition Lab at MIT for thoughtful discussion and feedback. Thanks also to the Discovery Center at the Museum of Science in Boston, the Children's Museum of Boston, and participating families.

References

Adolph, K., Eppler, M., & Gibson, E. (1993). Crawling versus walking infants' perception of affordances for locomotion over sloping surfaces. *Child Development, 64*(4), 1158–1174.

Berger, S., Adolph, K., & Lobo, S. (2005). Out of the toolbox: Toddlers differentiate wobbly and wooden handrails. *Child Development, 76*(6), 1294–1307.

Brauer, J., Kaminski, J., Reidel, J., Call, M., & Tomasello, M. (2006). Making inferences about the location of hidden food: Social dog, causal ape. *Journal of Comparative Psychology, 120*(1), 38–47.

Brown, A. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science: A Multidisciplinary Journal, 14*(1), 107–133.

Carey, S. (1987). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development, 1*(1), 37–41.

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711.

Daw, N., O'Doherty, J., Dayan, P., Seymour, B., & Dolan, R. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879.

Emery, N., & Clayton, N. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science, 306*, 1903–1907.

Gelman, S. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gibson, J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Gopnik, A., & Schulz, L. E. (Eds.). (2007). *Causal learning: Psychology, philosophy and computation*. New York: 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 patters of variation and covariation. *Developmental Psychology, 37*(5), 620–629.

Gweon, H., & Schulz, L. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers' exploratory play. *Paper presented at the proceedings of the 30th annual conference of the cognitive science society*. Washington, DC.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcome of physical events. *Child Development, 71*, 1540–1544.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic books.

Kaelbling, L. (1993). *Learning in embedded systems*. Cambridge, MA: The MIT Press.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., & Wang, J. T. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973.

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

Keil, F. C. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Sciences, 7*(8), 368–373.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661–667.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674–689.

Kuhn, D., Amsel, E., & O'Laughlin, M. (1988). *The development of scientific thinking skills*. New York: Academic Press.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science, 16*(9), 678–683.

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

Lockman, J. (2000). A perception–action perspective on tool use development. *Child Development, 71*(1), 137–144.

Masnick, A., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*(1), 67–98.

McClure, S., Daw, N., & Read Montague, P. (2003). A computational substrate for incentive salience. *Trends in Neurosciences, 26*(8), 423–428.

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: Orangutans use water as a tool. *Biology Letters, 3*(5), 453–455.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions, 6*(3), 38–43.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608–631.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521–562.

Schulz, L., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40*, 162–176.

Schulz, L., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*(3), 322.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43*(5), 1124–1130.

Shannon, C., & Weaver, W. (1948). *A mathematical theory of communication*. Champaign, IL: University of Illinois Press.

Sobel, D., & Munro, S. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology, 45*(2), 511–524.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303–333.

Stulp, G., Emery, N. J., Verhulst, S., & Clayton, N. S. (2009). Western scrub-jays conceal auditory information when competitors can hear but cannot see. *Biological letters, 5*, 583–585.

Sutton, R., & Barto, A. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

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

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and machine*. Cambridge, MA: MIT Press.