Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence

Daphna Buchsbaum [a,*], Alison Gopnik [a], Thomas L. Griffiths [a], Patrick Shafto [b]

[a] Department of Psychology, University of California, Berkeley, United States
[b] Department of Psychology, University of Louisville, United States

ABSTRACT

Children are ubiquitous imitators, but how do they decide which actions to imitate? One possibility is that children rationally combine multiple sources of information about which actions are necessary to cause a particular outcome. For instance, children might learn from contingencies between action sequences and outcomes across repeated demonstrations, and they might also use information about the actor's knowledge state and pedagogical intentions. We define a Bayesian model that predicts children will decide whether to imitate part or all of an action sequence based on both the pattern of statistical evidence and the demonstrator's pedagogical stance. To test this prediction, we conducted an experiment in which preschool children watched an experimenter repeatedly perform sequences of varying actions followed by an outcome. Children's imitation of sequences that produced the outcome increased, in some cases resulting in production of shorter sequences of actions that the children had never seen performed in isolation. A second experiment established that children interpret the same statistical evidence differently when it comes from a knowledgeable teacher versus a naïve demonstrator. In particular, in the pedagogical case children are more likely to "overimitate" by reproducing the entire demonstrated sequence. This behavior is consistent with our model's predictions, and suggests that children attend to both statistical and pedagogical evidence in deciding which actions to imitate, rather than obligately imitating successful action sequences.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Learning the causal relationships between everyday sequences of actions and their outcomes is a daunting task. How do you transform a package of bread, a jar of peanut butter and a jar of jelly into a peanut butter and jelly sandwich? Do you cut the bread in half before or after you put together the sandwich? Can you put the jelly on first, or does it always have to be peanut butter first? In order to achieve desired outcomes – from everyday goals such as eating a tasty sandwich, to complex tasks such as

making and using tools – children need to solve a challenging causal learning problem: observing that the intentional actions of others lead to outcomes, inferring the causal relations between actions and outcomes, and then using that knowledge to plan their own actions.

To learn from observation in this way, children cannot simply mimic everything they see. Instead, they must segment action sequences into meaningful subsequences, and determine which sequences are relevant to outcomes and why. Recent studies of imitation have produced varying answers to the question of whether children are capable of solving this problem. While children sometimes selectively reproduce the most obviously causally effective actions (Schulz, Hooppell, & Jenkins, 2008; Williamson, Meltzoff, & Markman, 2008), at other times they will "overimitate", reproducing apparently unnecessary parts

* Corresponding author. Address: University of California, Department of Psychology, 3210 Tolman Hall # 1650, Berkeley, CA 94720-1650, United States. Tel.: +1 510 643 2172; fax: +1 510 642 5293.

E-mail address: daphnab@berkeley.edu (D. Buchsbaum).

of a causal sequence (Lyons, Young, & Keil, 2007; Whiten, Custance, Gomez, Teixidor, & Bard, 1996), or copying an actor's precise behavior, when a more efficient action for accomplishing the goal is available (Meltzoff, 1995). Sometimes children may do both in the same study. In the "rational imitation" studies by Gergely, Bekkering, and Kiraly (2002), children saw an experimenter activate a machine with hands free or hands confined. Children both produced exact imitations of the actor (touching their head to a machine to make it go) and produced more obviously causally effective actions (touching the machine with a hand), though the proportion of such actions differed in the different intentional contexts. The evidence on children's use of intentional and pedagogical cues to inform their imitation is similarly varied, with studies showing that in some contexts children use information about the demonstrator's intentional and knowledge state to aid their causal inferences (Brugger, Lariviere, Mumme, & Bushnell, 2007; Williamson et al., 2008), while in others these cues can lead children astray (Bonawitz et al., this issue; Sobel & Sommerville, 2009).

We suggest that these different results reflect the multiple sources of information that contribute to a rational statistical inference about causally effective action sequences. Children need to balance prior knowledge about causal relations, the new evidence that is presented to them by the adult, and knowledge of the adult's intentions. Moreover, there is often no single "right answer" to the question of what to imitate. After all, a longer "overimitation" sequence might actually be necessary to bring about an effect, though that might initially seem unlikely.

Probabilistic models are well suited to combining multiple sources of information. In particular, the imitation problem can be expressed as a problem of Bayesian inference, with Bayes' rule indicating how children might combine these factors to formulate different causal hypotheses and produce different action sequences based on those hypotheses. It is difficult to test this idea however, without knowing the strength of various causal hypotheses for the children. Since previous studies involved general folk physical and psychological knowledge (such as removing a visibly ineffectual bolt to open a puzzle box) it is difficult to know how strong those hypotheses would be. By giving children statistical information supporting different hypotheses we can normatively determine how probable different hypotheses should be, and then see whether children's imitation reflects those probabilities.

It is also independently interesting to explore the role of statistical information in imitation. Recent studies show that children are surprisingly sophisticated in their use of statistical information such as conditional probabilities in a range of domains, from phonology (Saffran, Aslin, & Newport, 1996), to visual perception (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002), to word meaning (Xu & Tenenbaum, 2007). Such information plays a particularly important role in both action processing (Baldwin, Andersson, Saffran, & Meyer, 2008; Buchsbaum, Griffiths, Gopnik, & Baldwin, 2009; Swallow & Zacks, 2008) and causal inference (Gopnik et al., 2004; Gopnik & Schulz, 2007), and allows adults to identify causal subsequences within continuous streams of action (Buchsbaum et al., 2009).

Statistical inference might be particularly important to imitation because it could allow children to not only determine the causal relationship between action sequences and outcomes, but to identify irrelevant actions within causally effective sequences. Imagine that I am making a peanut butter sandwich, and that before opening the jar, I wipe my hands on a paper towel. If this is the first time you have seen me make a sandwich, you might mistakenly think that hand-wiping is a necessary step. However, after watching me make a sandwich a couple of times, you might notice that while I always turn the lid counter-clockwise before opening the jar, I do not always wipe my hands before opening the jar, and could infer that this step is extraneous. In most previous work on children's imitation of casual sequences, children were given only a single demonstration of how to generate the outcome (e.g. Lyons et al., 2007; Whiten et al., 1996).

In this paper, we first look at whether children use statistical evidence from repeated demonstrations to imitate the correct causal subsequence within a longer action sequence. We present a Bayesian analysis of causal inference from repeated action sequence demonstrations, followed by an experiment investigating children's imitative behavior and causal inferences. We showed preschool children different sequences of three actions followed by an effect, using our Bayesian model to guide our manipulation of the probabilistic evidence, such that the statistical relations between actions and outcomes differed across conditions in ways that supported different causal hypotheses. We then examine which sequences the children produced themselves, and compare children's performance to our model's predictions.

Second, we investigate whether children can combine pedagogical and knowledge state information with directly observed statistical evidence, to guide their imitative choices. Will children's behavior change as the learning context becomes more pedagogical? We compare children's imitative choices when observing a knowledgeable teacher versus a naïve demonstrator performing the same set of action sequences and outcomes. Children might assume that all adults, naïve or knowledgeable, are demonstrating potentially relevant actions, but the intuitive prediction is that children would be more likely to "overimitate" – reproducing every detail of the experimenter's actions – when the demonstrator is a knowledgeable teacher. We show how this intuition can be captured formally. We present an extension of our Bayesian model that makes behavioral predictions based on both information about statistics and about the demonstrator's knowledge, and compare children's performance to our model's predictions.

2. Bayesian ideal observer model

While it is intuitively plausible that children use statistical evidence from repeated demonstrations to infer causal structure, we would like to verify that normative inferences from repeated observations of action sequences and their outcomes vary in a systematic way with different patterns of data. One way to derive what the normative distribution over causes should be is through a Bayesian

model (Gopnik et al., 2004; Griffiths & Tenenbaum, 2005). The Bayesian formalism provides a natural way for us to explicitly represent the roles of both children's prior knowledge, and the observed data in forming children's beliefs about which action sequences are likely to be causal.

2.1. Model details

Given observations of several action sequences, we assume that children consider all sequences and terminal subsequences as potentially causal. For instance, if the sequence "squeeze toy, knock on toy, pull toy's handle" is observed, then squeeze, followed by knock, followed by pull handle would be one possible causal sequence, and knock followed by pull handle would be another. Given all of the observed sequences, we can enumerate the potential causes (see Table 1 for an example set of demonstrations and potential causes). As in previous work on children's causal inference, we use a deterministic-OR model (Griffiths & Tenenbaum, 2009), in which any of the correct sequences will always bring about the effect. To capture the intuition that there may be multiple action sequences that bring about an effect, we consider combinations of up to five individual causal sequences. A hypothesis, h, represents one possible combination of causal sequences, and the hypothesis space H contains all such possible combinations (see Fig. 1).

Table 1
Example demonstrations, and the associated set of potential causal sequences.

Observed action sequence	Potential causal sequences
ABC+	ABC, BC, C
DBC+	DBC, BC, C
Total Potential Causes	ABC, DBC, BC, C

Note: Letters represent unique observed actions (e.g. A = Knock, B = Roll, C = Squish) while a + indicates a causal outcome.

From the learner's perspective, the problem is that they observe an action sequence, and then observe whether or not the effect is elicited. Based on this information, they want to infer what sequences of actions cause the effect. More formally, the learner wants to infer the set of causal sequences, h, given the observed data, d, where the data are composed of an observed action sequence, a, and an outcome, e. Bayes' theorem provides a way to formalize this inference. Bayes' theorem relates a learner's beliefs before observing the data, their prior $p(h)$, to their beliefs after having observed the data, their posterior $p(h-d)$,

$$p(h|d) \propto p(d|h)p(h), \tag{1}$$

where $p(d|h)$ is the probability of observing the data given the hypothesis is true. For deterministic-OR causal models, this value is 1 if the sequence is consistent with the hypothesis, and zero otherwise. For example, given the hypothesis that squeeze is the cause, a consistent observation would be, knock then squeeze followed by music, and an inconsistent observation would be squeeze followed by no music. When multiple sequences of actions and effects are observed, we assume that these sequences are independent.

A key element in this inference is the learner's prior expectations, $p(h)$. Previous research suggests that children believe there tends to be only one correct sequence, as opposed to many possible sequences, that cause an effect (e.g. Sobel, Tenenbaum, & Gopnik, 2004). It also suggests that, all else being equal, children believe adults to be rational actors who do not perform extraneous actions (e.g. Gergely et al., 2002). We capture these intuitions with a prior that depends on two parameters, p and β, which correspond to the learner's expectations about the number of ways to generate an effect, and about the length (in actions) of causal sequences. We might say that p reflects the strength of children's simplicity bias, while β represents the degree to which they believe adults will not produce irrelevant actions (thus leading the children to think that longer subsequences of the adult demonstrations are more
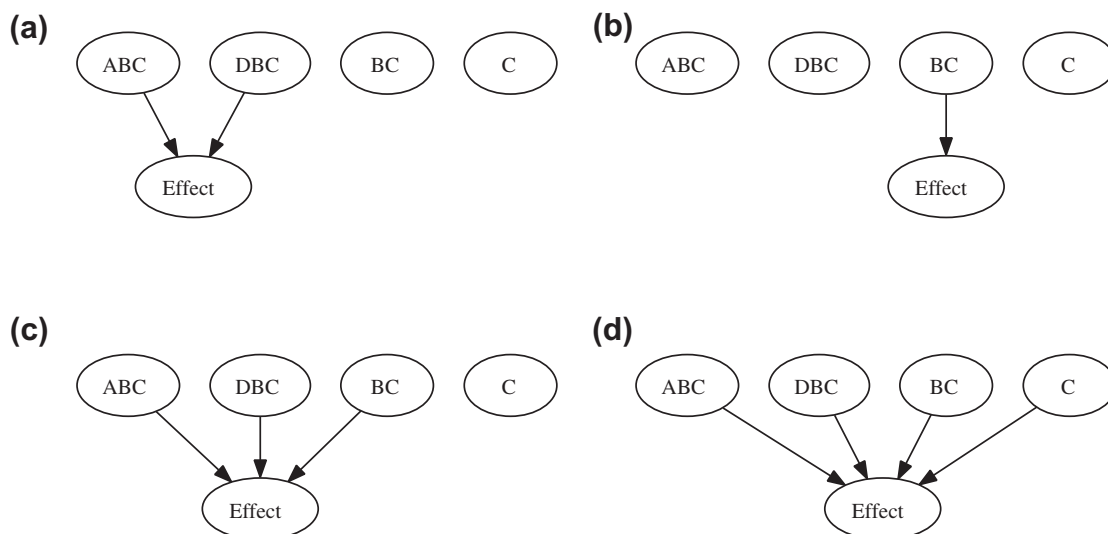

Fig. 1. Part of an example hypothesis space. Graphs (a)–(d) each represent a different hypothesis about which action sequences are causal.

likely to be causal). Note that these two assumptions may be in tension and so the model (and the children) will have to balance them.

We formalize the prior as a generative model, where hypotheses are constructed by randomly choosing causal sequences, a. Each sequence has a probability p_a of being included in each hypothesis and a probability $(1 - p_a)$ of not being included,

$$p(h) \propto \prod_{a \in h} p_a \prod_{a* \notin h} (1 - p_{a*}) \qquad (2)$$

where the probability of including causal sequence a is

$$p_a = \frac{1}{1 + \frac{1-p}{p} \exp(-\beta(|a| - 2))} \qquad (3)$$

and $|a|$ is the number of actions in the sequence a. Values of β that are greater than 0 represent a belief that longer sequences are more likely to be causes. Values of p less than 0.5 represent a belief that effects tend to have few causal sequences. Taken together, Eqs. (1)–(3) provide a model of inferring hypotheses about causes from observed sequences and their effects.

In our experiments, rather than probing children's beliefs directly, we allow children to play with the toy. Therefore, to complete the model, we must specify how children choose action sequences, a, based on their observations, d. Intuitively, we expect that if we know the set of causes of the effect, h, we will randomly choose one of these sequences. If we were unsure about which of several possible causes was the right one, then we may choose any of the possible contenders, but biased toward whichever one we thought was most likely. We capture these intuitions formally by choosing an action sequence given the observed data, $p(a|d)$, based on a weighted sum over possible hypotheses,

$$p(a|d) = \sum_{h \in H} p(a|h) p(h|d) \qquad (4)$$

where $p(a|h)$ is one over the number of causes consistent with h, $1/|h|$, and $p(h|d)$ is specified in Eq. (1). Causal models using similar probability matching have successfully predicted children and adult's performance on a variety of tasks (Griffiths & Tenenbaum, 2009).

2.2. A simple modeling example

We can now verify that the model makes distinct inferences from repeated demonstrations. In the first example, the demonstrated action sequences are ABC+, DBC+ as in Table 1. That is, a sequence of three actions A, B and C is followed by an effect. Subsequently, a different sequence of three actions D, B, and C is followed by the same effect. In the second example, the observed sequences are ABC+, DBC−. In this case, the second three-action sequence is not followed by the effect.

Using values of $p = 0.5$ and $\beta = 0$ results in a prior that assigns equal probability to all possible causal hypotheses – a uniform prior. With this uniform prior, our model infers that, in the first case, all the sequences are possible causes, with BC and C being somewhat more likely, and equally

Table 2
Example model results, $p = 0.5$ and $\beta = 0$.

Observed sequences	ABC	DBC	BC	C
ABC+, DBC+	0.21	0.21	0.28	0.28
ABC+, DBC−	1.0	0.0	0.0	0.0

Note: Values are the probability of choosing to perform this action sequence to bring about the effect given the observed data, $p(a|d)$, as described in Eq. (4).

Table 3
Example model results, $p = 0.1$ and $\beta = 1.4$.

Observed sequences	ABC	DBC	BC	C
ABC+, DBC+	0.28	0.28	0.34	0.09
ABC+, DBC−	1.0	0.0	0.0	0.0

Note: Values are the probability of choosing to perform this action sequence to bring about the effect given the observed data, $p(a|d)$, as described in Eq. (4).

probable. Notice that the model infers that the subsequences BC and C are the most likely causes, even though neither was observed on its own. The second case is quite different. Here the model sees that DBC and its subsequences BC and C did not lead to the effect in the second demonstration, and infers that ABC is the only possible cause among the candidate sequences (see Table 2).

We now use values of $p = 0.1$ and $\beta = 1.4$ leading the model to favor simpler hypotheses containing fewer causes, and causes that use more of the observed demonstration.[1] This prior does not change results in the second case, where ABC is still the only possible cause. However, in the first case, the model now infers that the subsequence BC is the most likely individual cause, since it is the longest observed sequence to consistently predict the effect (see Table 3).

2.3. Model predictions for children's inferences

We can now use the model to help us construct demonstration sequences that normatively predict selective imitation in some cases, and "overimitation" in others. If children are also making rational inferences from variations in the action sequences they observe, then their choice of which actions to imitate in order to bring about an effect should similarly vary with the evidence. We test our prediction that children rationally incorporate statistical evidence into their decisions to imitate only part of an action sequence versus the complete sequence in the following sections.

3. Experiment 1

3.1. Method

3.1.1. Participants
Participants were 81 children ($M = 54$ months, $Range = 41–70$ months, 46% female) recruited from local prescho-

---

[1] These parameter values are those that produce the best fit to children's imitation behavior in Experiment 1, as we discuss later in the paper.

ols and a science museum. Another 18 children were excluded from the study because of demonstration error (4), equipment failure (3), lack of English (1), unavailable birth date (1), did not try toy (6), extreme distraction (2), never performed trial termination action (1).

3.1.2. Stimuli

There were two novel toys: a blue ball with rubbery protuberances, and a stuffed toy with rings and tabs attached to it. Six possible actions could be demonstrated on each toy. Toys were counterbalanced across children. Children were assigned to one of three experimental conditions. In each condition, they saw a different pattern of evidence involving five sequences of action and their outcomes. Each individual action sequence was always three actions long. In the "ABC" pattern, the same sequence of three actions (e.g. A = Knock, B = Stretch, C = Roll) is followed by a musical effect three times, while in the "BC" pattern a sequence composed of a different first action, followed by the same two-action subsequence (e.g. A = Squish, B = Pull, C = Shake and D = Flip, B = Pull, C = Shake) is followed by the effect three times (see Table 4). In both patterns, two additional sequences that end in C and do not contain BC fail to produce the effect. Finally, in the "C" pattern the sequences of actions were identical to those in the "BC" pattern, but the outcome was always positive. The number of times each individual action is demonstrated in each sequence position is identical in all three patterns. As we show later in the paper, our Bayesian ideal observer model confirms that the statistical evidence in each pattern supports different causal inferences.

3.1.3. Procedure

The experimenter showed the child one of the toys, and said: "This is my new toy. I know it plays music, but I haven't played with it yet, so I don't know how to make it go. I thought we could try some things to see if we can figure out what makes it play music". The experimenter emphasized her lack of knowledge, so that the children would not assume she knew whether or not any of her actions were necessary. She then demonstrated one of the three patterns of evidence, repeating each three-action sequence (and its outcome) twice. The experimenter named the actions (e.g. "What if I try rolling it, and then shaking it, and then knocking on it?"), acted pleasantly surprised when the toy played music ("Yay! It played music'!"), or disappointed when it did not ("Oh. It didn't go"), and pointed out the outcome ("Did you hear that song?" or "I don't hear anything. Do you hear anything?"). After she demonstrated all five of the 3-action sequences, she gave the child the toy and said "Now it's your turn! why don't

you try and make it play music." Throughout the experiment the music was actually triggered by remote activation. To keep the activation criteria uniform across conditions, the toy always played music the first time a child produced the final C action, regardless of the actions preceding it, terminating the trial. Only this first sequence of actions was used in our analysis. Each child interacted with one toy, in a single condition of the experiment.

Children were videotaped, and their actions on the toy from the time they were handed the toy to trial termination were coded by the first author, and 80% of the data was recoded by a blind coder. Coders initially coded each individual action children performed as one of the six demonstrated actions, or as "novel". These sequences were then transferred into an "ABC" type representation, and subsequently coded as one of four sequence types: Triplet, Double, Single or Other (defined below). Inter-coder reliability was very high, with 91% agreement on the "ABC" type representations, and 100% agreement on sequence types.

3.2. Results and discussion

Children produced significantly different types of sequences across the three conditions, $p < 0.001$ (two-sided Fisher's exact test, Table 5). There was no difference in sequence types produced by children interacting with the two different toys ($p = 0.40$, n.s., two-sided Fisher's exact test). We will discuss results for the "ABC" and "BC" conditions first, and then return to the "C" condition.

3.2.1. Effect of statistical evidence on imitation

In their imitation, children could either exactly reproduce one of the three-action sequences that had caused the toy to activate (that is, ABC in the "ABC" condition or ABC, DBC or EBC in the "BC" condition), or they could just produce BC in isolation. We refer to these successful three-action sequences as "triplets", and to the BC subsequence as a "double".

Both a triplet and a double reflect potentially correct hypotheses about what caused the toy to activate in both conditions. It could be that BC by itself causes the toy to activate in the "ABC" condition and the A is superfluous, or it could be that three actions are necessary in the "BC" condition, but the first action can vary.

If children automatically encode the adult's successful actions as causally necessary, then they should exclusively imitate triplets in both conditions. However, if children are also using more complex statistical information, they should conclude that the BC sequence by itself is more likely to be causal in the "BC" condition than in the "ABC" condition, and that the triplet sequence is more likely to be causal in the "ABC" condition than in the

Table 4
The demonstration sequences for "ABC", "BC" and "C" conditions.

"ABC" condition	"BC" condition	"C" condition
ABC+	A**BC+**	AB**C+**
DEC	ADC	AD**C+**
ABC+	D**BC+**	DB**C+**
EDC	AEC	AE**C+**
ABC+	E**BC+**	EB**C+**

Table 5
Number of children producing each sequence type in each condition of Experiment 1.

Condition	Triplet	Double	Single	Other
"ABC"	20	1	2	4
"BC"	10	7	0	10
"C"	8	0	8	11

"BC" condition. This is in fact what we found – the number of children producing triplets and doubles varied by condition, $p < 0.01$ (two-sided Fisher's exact test, Table 5, columns 1 and 2), and differed significantly between the "ABC" and "BC" conditions $p < 0.05$ (two-sided Fisher's exact test, Table 5, columns 1 and 2, "ABC" and "BC" conditions).

3.2.2. Effect of differing causal outcomes on imitation

Children in the "BC" condition saw three different action sequences precede the effect, while children in the "ABC" condition saw only one sequence precede the effect. This may have confused children in the "BC" condition, leading them to produce a variety of random actions, including BC. The "C" condition controls for this possibility. In this condition the sequences of actions were identical to those in the "BC" condition, but the outcome was always positive. As we show later, our Bayesian ideal observer model confirms that this provided statistical evidence for the hypothesis that C alone was sufficient to produce the effect.

In all three conditions, imitation of just the final C action in isolation was coded as a "single". As in the "ABC" and "BC" conditions, only the subsequence BC was coded as a double in the "C" condition. Also consistent with the "ABC" and "BC" conditions, in the "C" condition all five demonstrated successful sequences (ABC, ADC, DBC, AEC and EBC) were coded as triplets.

The "C" condition is as complex as the "BC" condition. However in the "C" condition the final action C produced by itself reflects a likely causal hypothesis. If children selectively imitate subsequences based on the data, then children in the "C" condition should produce C more frequently than children in the "BC" condition, and children in the "BC" condition should produce BC more frequently than children in the "C" condition. Our results support this hypothesis. Children in the "BC" and "C" conditions differed significantly in the overall types of sequences they produced, $p < 0.001$ (two-sided Fisher's exact test, Table 5 "BC" condition and "C" condition), and the number of children producing doubles and singles in the two conditions also varied significantly, $p < 0.001$, (two-sided Fisher's exact test, Table 5, columns 2 and 3, "BC" and "C" conditions).

Finally, a split by median age (*Median* = 56 months), revealed no differences in performance between older and younger age groups for any of the above analyses (two-sided Fisher's exact tests, Table 6), consistent with previous results with this age range (Lyons et al., 2007; McGuigan, Whiten, Flynn, & Horner, 2007).

3.2.3. Performance of "Other" actions

Across all conditions, children did not just obligately imitate one of the successful sequences or subsequences

Table 6
Number of children producing each sequence type in Experiment 1, median split by age.

Condition	Triplet	Double	Single	Other
Older	19	6	4	13
Younger	19	2	6	12

they observed – they also produced new combinations of actions. Overall, the types of "Other" sequences produced did not qualitatively differ across conditions, and appear to be a mix of exploratory behavior (e.g. performing the sequence BEC in the "BC" condition or BABC in the "ABC" condition) and genuine errors (e.g. producing ADC in the "BC" condition). There was a trend towards children in the "BC" and "C" conditions performing more of these "Other" sequences than children in the "ABC" condition $p = 0.10$, (two-sided Fisher's exact test). This difference becomes statistically significant when the two children who imitated unsuccessful triplets (e.g. ADC) are excluded from the analysis, leaving only children who performed sequences they had never seen, and subsequences other than BC and C (DC, AC or EC) $p < 0.05$, (two-sided Fisher's exact test). This result is compatible with findings that children increase their exploratory behavior when the correct causal structure is ambiguous (Schulz & Bonawitz, 2007; Schulz et al., 2008). Finally, four children, all in the "BC" and "C" conditions, performed novel actions (e.g. throwing the ball) or actions they had never seen demonstrated, consistent with these conditions eliciting more exploratory actions.

4. Modeling Experiment 1

Consistent with our experimental results, our model makes distinct predictions in each of the three experimental conditions, showing that the data supports differential causal inferences. However, we would like to explore the quantitative predictions of the model in a bit more detail.

Recall that our model has two parameters, β and p, which correspond to the learner's pre-existing expectations about the length of causal sequences and number of ways to generate an effect. By fitting the model parameters to the behavioral data from Experiment 1, we can not only evaluate the model predictions more quantitatively, we can also determine the nature and strength of these same assumptions for children.

Model fit was determined by measuring the distance between the model predictions and the observed data. Because solving for the best fitting parameters is not analytically tractable, we used a grid search over the range [0, 1] for p and [0, 2] for β to find the best fitting parameters. While the qualitative (and quantitative) fit of the model was robust across a range of parameters, we found that the parameters $p = 0.1$ and $\beta = 1.4$ provided the best quantitative fit to the data from Experiment 1. These parameter values minimize both sum of squared error ($SSE = 0.115$) and χ^2 distance ($\chi^2 = 0.068$). These values are used throughout this paper, allowing a generalization test of the model predictions in Experiment 2.

We used Pearson's correlation coefficient, $r = 0.93$, as a measure of the model's fit to the data. This close match to children's performances (see Fig. 2) suggests that children's inferences based on the naïve demonstrator's actions conform closely to normative predictions based on the demonstrated action sequences. It also suggests that children may be considering the probability of several hypotheses rather than simply settling on one hypothesis and eliminating the rest.

(a)

Proportion of Children Performing Sequence Type

(b)

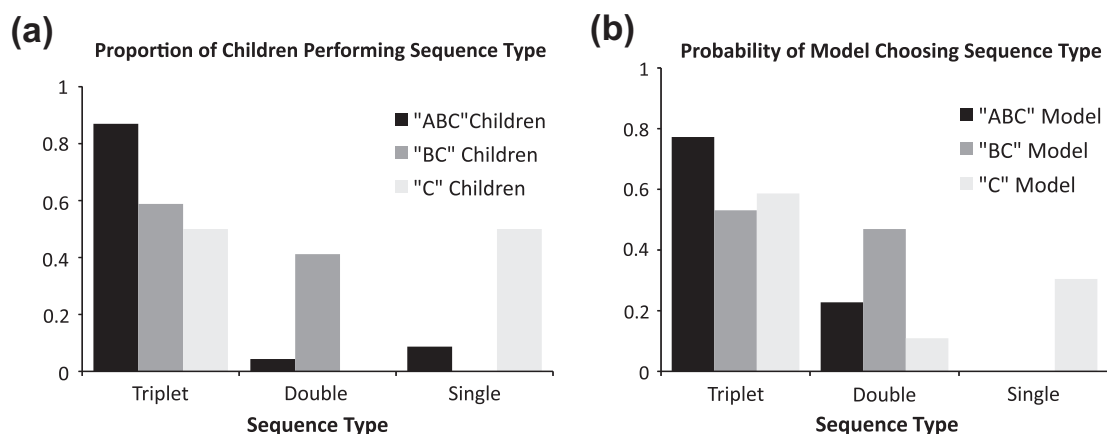**Probability of Model Choosing Sequence Type**

Fig. 2. Modeling the results of Experiment 1. (a) Children's performance. (b) Predictions of our Bayesian model.

Finally, the relatively low value for *p* suggests that children employ a causal Ockham's razor, assuming that simpler hypotheses, which require fewer causal sequences to explain the data, are more likely than more complex hypotheses. The relatively high value for *β* in the best fitting model suggests that children prefer individual causal sequences to use more of the demonstrated actions, perhaps representing a pre-existing belief that, as rational actors, adults usually do not perform extraneous actions.

Children might make this "rational actor" assumption because they are using information about the knowledge-ability (e.g. Jaswal & Malone, 2007), reliability (e.g. Koenig, Clement, & Harris, 2004; Zmyj, Buttelmann, Carpenter, & Daum, 2010) and intentional stance (e.g. Bonawitz et al., this issue) of the demonstrator. For instance, children might notice that the experimenter always performs three-action sequences, and infer that the experimenter, while not knowing the correct sequence, knows that it must be three actions long. We next present an extension of our model that explicitly incorporates stronger pedagogical and knowledge state information, in addition to statistical evidence.

5. Learning from knowledgeable pedagogical demonstrators

Children may learn from observing individuals who don't know how a toy works, as in Experiment 1, or they may learn from a helpful teacher who is choosing examples to try to teach the child how the toy works. In teaching situations, children may draw different inferences from the same data by inferring *why* the teacher chose these data. Intuitively, children may implicitly assume that the teacher's sample demonstrations are not randomly chosen, but are designed to be informative (Csibra & Gergely, 2006).

We can formalize this idea by incorporating a model of how a teacher's choice of interventions provides information about the hypothesis they are trying to teach into our initial model of rational imitation. We can then compare our model's predictions to children's performance,

to see if children's imitative choices reflect a belief that knowledgeable teachers select informative examples.

5.1. Modeling pedagogical learning

Recall Eq. (1) related a learner's posterior beliefs $p(h|d)$ to their prior beliefs, $p(h)$. This was accomplished by way of a measure of how consistent the data were with a hypothesis, $p(d|h)$. Here, the data, *d*, include an action sequence, *a*, and an outcome *e*. We did not specify our belief about how the demonstrator's sequence of actions, *a*, was chosen. Implicitly, we assumed that these choices were random, and therefore did not factor into our inference. However, to formalize how having a helpful teacher may affect inferences, we must specify how the demonstrator chooses their actions and expand Eq. (1) to include a factor, $p(a|h)$. The learner would then update their beliefs based on the product of the prior probability, the probability of the action given a hypothesis, and the probability of the effect given the action and the hypothesis

$$p(h|a, e) \propto p(e|h, a)p(a|h)p(h) \qquad (5)$$

Here we have introduced $p(a|h)$, which specifies the learner's beliefs about how the demonstrator chooses their action sequence given a hypothesis, and separated the data into the action sequence, *a*, and it's effects, *e*. For a demonstrator who was choosing their actions at random, $p(a|h)$, is the same for all sequences, $\frac{1}{|A|}$ (where *A* is the set of all action sequences, and $|A|$ is the number of possible sequences) and can be ignored. However, if the learner believes the demonstrator is a helpful teacher, then they could expect the teacher to choose their actions, $p(a|h)$, with the goal of having the learner infer the correct hypothesis,

$$p_t(a|h) \propto p_l(h|a, e) \qquad (6)$$

where *t* and *l* indicate teacher and learner, respectively (Shafto & Goodman, 2008). The equation states that the learner can expect the teacher to choose action sequences that tend to make the learner believe the correct hypothesis.

5.2. Model predictions

By explicitly representing assumptions about the demonstrator's knowledgeability and helpfulness, the pedagogical model makes distinctly different predictions than the previous model. The pedagogical model assumes that the demonstrator has not chosen their actions randomly, but for the purpose of teaching the learner. This implies that the learner should put more weight in the demonstrations, as compared to the same evidence demonstrated by a naïve individual. Therefore, if the teacher chose to demonstrate a long sequence such as squish, knock, pull and the effect was elicited, the learner would be more likely to infer that all three actions were necessary, than if these demonstrations were produced randomly (for other work on pedagogical inference, see Shafto & Goodman (2008) & Bonawitz et al. (this issue)).

Consider the BC condition from Experiment 1 (see Table 4). Children observed five sequences of actions, three of which led to the effect and two that did not. Of the three cases that elicited the effect, all contained the subsequence BC, and when the effect was not elicited this subsequence was not present. However, in all of the sequences, the demonstrator chose sequences of three actions. Under the assumption that the demonstrator is naïve, the model predicted that these factors trade-off, leading to the prediction that it is roughly equally likely that triplets or doubles could elicit the effect.

In contrast, under the assumption that the demonstrator is knowledgeable and helpful, the pedagogical model predicts a shift in children's inferences. Fig. 3 shows the predictions of the model assuming naïve and pedagogical demonstrators (and the parameter values used in the first experiment). The pedagogical model predicts that, after observing the same sequences of actions, children should be much more inclined to believe that triplets cause the effect. We test this prediction in the following experiment.
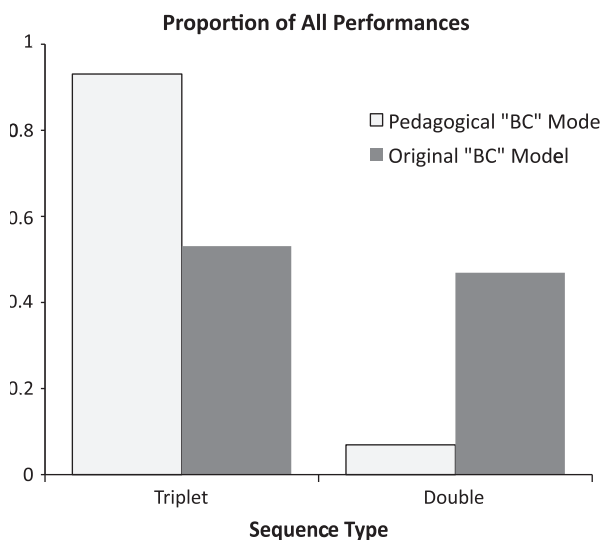

Fig. 3. Predictions of our model given assumptions of pedagogical sampling (as in Experiment 2) or random sampling (as in Experiment 1).

6. Experiment 2: Effect of combined pedagogical and statistical evidence on imitation

6.1. Method

6.1.1. Participants

Twenty seven children (M = 52 months, *Range* = 44–62 months, 37% female) recruited from preschools and a science museum were included in this study. Another 11 children were excluded because of experimenter error (4), equipment failure (1), parental interference (1), extreme distraction (1), never performed trial termination action (1), failure to complete experiment (3).

6.1.2. Stimuli

The same two novel toys and corresponding actions were used as in Experiment 1. In this condition, the demonstrated sequences of actions and outcomes were identical to those in the "BC" condition of Experiment 1.

6.1.3. Procedure

The experimenter showed the child one of the toys, and said: "See this toy? This is my toy, and it plays music. I'm going to show you how it works. I'll show you some things that make it play music and some things that don't make it play music, so you can see how it works". The experimenter emphasized her knowledge of the toy, and that her actions were chosen purposefully and pedagogically. She then demonstrated the "BC" pattern of evidence, almost exactly as in the BC condition of Experiment 1. The only difference was that the experimenter indicated that she expected each resulting outcome. ("See? It played music" or "See? No music".) Otherwise the procedure and coding was exactly as in Experiment 1. Inter-coder reliability was very high, with 91% agreement on the "ABC" type representations, and 100% agreement on sequence types.

6.2. Results and discussion

The action sequences and causal relationships demonstrated in this experiment are identical to those in the "BC" condition of Experiment 1. If children are only attending to the observed statistical evidence, then their inferences here should be the same as in the original "BC" condition. However, since children are now told that the experimenter is showing them how the toy works, this explicit pedagogy provides additional causal information. If children believe that the demonstrator is a rational teacher, then they might think that the demonstrator is choosing to show them triplets, because triplets, not doubles, are necessary to produce the effect, and should shift their imitative choices accordingly. Therefore, if children are able to attend to both statistical evidence and the demonstrator's pedagogical stance, then they should produce more triplets in the pedagogical "BC" condition than the original "BC" condition, and more doubles in the original "BC" condition than in the pedagogical "BC" condition.

Children in the original and pedagogical "BC" conditions differed significantly in the types of sequences they produced, $p < 0.05$ (two-sided Fisher's exact test, Table 7).

Table 7
Number of children producing each sequence type in Experiment 2.

Condition	Triplet	Double	Single	Other
Naïve "BC"	10	7	0	10
Pedagogical"BC"	14	0	0	13

The number of doubles and triplets produced in the two conditions varied significantly, $p < 0.01$, (two-sided Fisher's exact test, columns 2 and 3, Table 7). As in Experiment 1, there was no difference in sequence types produced by children interacting with the two different toys ($p = 0.70$, n.s., two-sided Fisher's exact test), and a split by median age (Median = 52 months) revealed no difference in sequence types produced by younger versus older children ($p = 0.45$, n.s., two-sided Fisher's exact test)

We used Pearson's correlation coefficient, $r = 0.99$, as a measure of the model's fit to the data (see Fig. 4). This close match to children's performances was achieved with the same parameters as were used in Experiment 1. This provides evidence that the complexity of the model is comparable to that of children's behavior, as we would expect an overly complex model to overfit the data and generalize poorly. Psychologically, these results suggest that children's inferences based on observations of a naïve demonstrator versus a knowledgeable teacher conform closely to normative predictions.

7. General discussion

In this paper, we examined whether children are sensitive to multiple sources of causal information when choosing the actions they imitate, and can integrate this information rationally. In Experiment 1, we demonstrated that children can use statistical evidence to decide whether to imitate a complete action sequence, or to selectively imitate only a subsequence. In particular, children in the "ABC" condition imitated the complete sequence ABC more often than children in the "BC" condition, while children in the "BC" condition imitated the subsequence BC more often than children in the "ABC" condition. Children's performance in the "C" condition demonstrated that the

differential imitation in the "ABC" and "BC" conditions could not be explained as a result of task complexity. In Experiment 2, we showed that children can combine statistical evidence with information about the demonstrator's knowledge state in deciding which actions to imitate – imitating different portions of the same action sequences when they observe them being performed by a helpful teacher versus a naïve demonstrator.

These results extend earlier findings that show children take causal and intentional information into account appropriately in their imitation. They show that children also take into account statistical information about the conditional probability of events and do so in an at least roughly normative way. Both the model and data suggest that children may be making more finely-graded judgments about the probability of various options rather than simply making yes or no decisions about whether to use a particular strategy. However, it should be pointed out that we had only one response per child in this study so that we do not know for sure whether this probability matching behavior applies to individual children or only to children as a group (for a discussion of probability matching behavior see for example Vulkan (2000) & Denison et al. (2009)).

The studies also suggest a rational mechanism for the phenomenon of "overimitation" (Lyons et al., 2007). In particular, the "triplet" responses could be thought of as a kind of overimitation, reproducing parts of a causal sequence that are not actually demonstrably necessary for the effect. These results suggest that this behavior varies depending on the statistics of the data and the probability of various hypotheses concerning them.

"Overimitation" also varies depending on the pedagogical intentions of the demonstrator. Our naïve demonstrator explicitly established her lack of knowledge. In contrast, the earlier studies of imitation we outlined at the start of this paper did not provide the child with either clearly pedagogical or non-pedagogical demonstrators.

These demonstrators may have used cues such as directed gaze and pointing (Csibra & Gergely, 2006), leading children to assume the demonstrated sequences were pedagogically sampled. In general, these studies also only provided children with a single demonstration, and used causal systems where children's prior expectations were
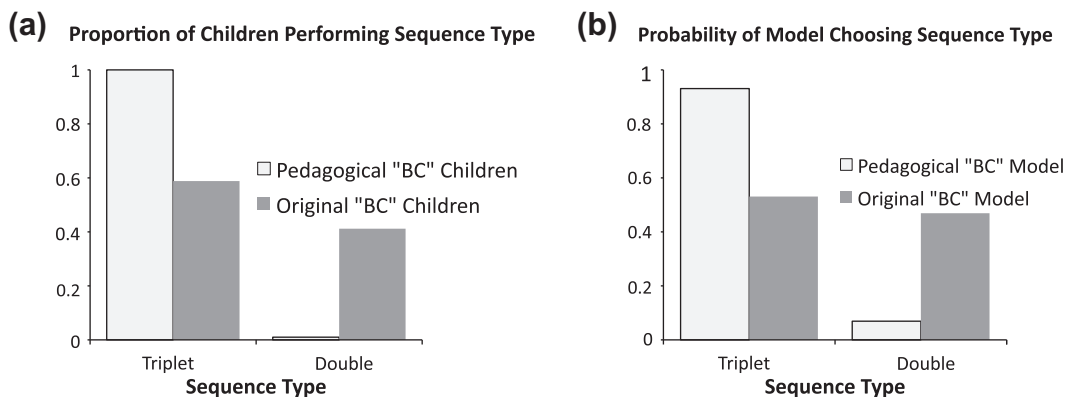

Fig. 4. Modeling the results of Experiment 2, using assumptions of Pedagogical sampling. (a) Children's performance. (b) Our model's predictions.

unknown. These differences may help explain the variance in outcomes across studies. This is the first study showing that children are more likely to overimitate when exactly the same actions are presented in an explicitly pedagogical versus non-pedagogical context. The model also suggests however, that despite appearances, such behavior is a rational response to a knowledgeable pedagogical demonstrator.

A related possibility, which we have not yet investigated empirically, is that seeing a repeated sequence of actions with no obvious physical causal outcome may lead children to suspect that the actions are intended to have a social or psychological rather than physical effect. Such inferences could be responsible for the use of imitation to transmit cultural conventions such as manners, rituals or even linguistic regularities.

These studies show that children are sensitive to statistical information, knowledge state, and pedagogical intention in determining which sequences of actions to imitate. Along with other studies, they suggest that Bayesian inference, which supports the construction of causal models from statistical patterns, may play a significant role in many important kinds of early learning. From learning how to make peanut butter sandwiches to playing with a new toy, children flexibly make use of many sources of information to understand the causal structure of the world around them.

Acknowledgements

References

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*, 1382–1407.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N., Spelke, E. S., & Schulz, L. E. (this issue). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development, 78*(3), 806–824.

Buchsbaum, D., Griffiths, T. L., Gopnik, A., & Baldwin, D. (2009). Learning from actions and their consequences: Inferring causal variables from continuous sequences of human action. In *Proceedings of the 31st annual conference of the Cognitive Science Society*.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of Change in Brain and Cognitive Development. Attention and Performance, 21*, 249–274.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2009). Preschoolers sample from probability distributions. In *Proceedings of the 32nd annual conference of the Cognitive Science Society*.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structures from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 458–467.

Gergely, G., Bekkering, H., & Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1–31.

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

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review, 116*(4), 661–716.

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-Year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*(3), 263–283.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition, 83*, B35–B42.

Koenig, M. A., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*(10), 694–698.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences, 104*, 19751–19756.

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*(3), 353–364.

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

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

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., Hooppell, C., & Jenkins, A. C. (2008). Judicious imitation: Children differentially imitate deterministically and probabilistically effective actions. *Child Development, 79*(2), 395–410.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the 30th annual conference of the Cognitive Science Society*.

Sobel, D. M., & Sommerville, J. (2009). Rationales in children's causal learning from others' actions. *Cognitive Development, 24*(1), 70–79.

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.

Swallow, K. M., & Zacks, J. M. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review, 15*(1), 116–122.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*, 101–118.

Whiten, A., Custance, D. M., Gomez, J. C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*(1), 3–14.

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

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review, 114*, 2.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology, 106*, 208–220.