Cognition 133 (2014) 343–357

Contents lists available at ScienceDirect

Cognition

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

Explaining prompts children to privilege inductively rich properties

Caren M. Walker [a,*], Tania Lombrozo [a], Cristine H. Legare [b], Alison Gopnik [a]

[a] University of California, Berkeley, Department of Psychology, United States
[b] University of Texas at Austin, Department of Psychology, United States

ARTICLE INFO

Article history:
Received 4 June 2013
Revised 19 July 2014
Accepted 22 July 2014

Keywords:
Explanation
Causal reasoning
Category labels
Non-obvious properties
Inductive inference
Generalization

ABSTRACT

Four experiments with preschool-aged children test the hypothesis that engaging in explanation promotes inductive reasoning on the basis of shared causal properties as opposed to salient (but superficial) perceptual properties. In Experiments 1a and 1b, 3- to 5-year-old children prompted to explain during a causal learning task were more likely to override a tendency to generalize according to perceptual similarity and instead extend an internal feature to an object that shared a causal property. Experiment 2 replicated this effect of explanation in a case of label extension (i.e., categorization). Experiment 3 demonstrated that explanation improves memory for clusters of causally relevant (non-perceptual) features, but impairs memory for superficial (perceptual) features, providing evidence that effects of explanation are selective in scope and apply to memory as well as inference. In sum, our data support the proposal that engaging in explanation influences children's reasoning by privileging inductively rich, causal properties.

1. Introduction

The challenge of causal reasoning is to discover the underlying structure of the world to facilitate prediction and action. This is non-trivial task. Despite the often strong correlation between what an object looks like and its causal properties (see Gelman & Medin, 1993), it is not uncommon to observe dissociations. In fact, perceptually similar objects can be endowed with very different causal properties: Poison hemlock may *look* identical to wild carrot, but it is certainly not good to eat. Learning how and when to override perceptual properties as a basis for judgment and action, and to instead favor inductively rich properties (such as causal affordances), is thus an important step in cognitive development.

We propose that the process of seeking, generating, and evaluating explanations plays an important role in encouraging children to recognize and privilege inductively-rich properties as a basis for reasoning, even when those properties are not perceptually salient. In particular, engaging in explanation could help children appreciate causal properties and subtle but reliable cues to causal structure, such as internal parts and category membership. For example, trying to explain why consuming hemlock generates one outcome (namely death) while consuming wild carrots generates another (perhaps pleasure) could help children appreciate that each plant has important internal properties, and that these internal properties are correlated with causal consequences they may wish to prevent (e.g., death) or to predict (e.g., pleasure).

In what follows, we first outline our proposal for the effects of explanation, motivating our hypothesis that explaining leads children to privilege inductively rich properties (i.e., those that facilitate a broad set of useful inferences). We then provide a brief review of prior

---

* Corresponding author. Address: University of California, Berkeley, Department of Psychology, 3210 Tolman Hall, Berkeley, CA 94720, United States.

E-mail address: caren.walker@berkeley.edu (C.M. Walker).

http://dx.doi.org/10.1016/j.cognition.2014.07.008

research on children's inductive generalizations in tasks that require choosing between a salient perceptual property (e.g., an object's color and shape) and a causal property (e.g., activating a machine). This body of research helps lay out the methods and developmental changes that motivate the current experiments.

1.1. Explanation and inference

Accounts of explanation from both philosophy and psychology suggest that explaining past and present observations can foster the acquisition of information that supports future actions and predictions (e.g., Craik, 1943; Friedman, 1974; Gopnik, 2000; Heider, 1958; Kitcher, 1989; Lombrozo, 2012; Lombrozo & Carey, 2006; Walker, Lombrozo, Williams, & Gopnik, submitted for publication; Walker, Williams, Lombrozo, & Gopnik, 2012). These ideas about the *functions or consequences* of explanation are consistent with several accounts of the *form and content* of explanations. In particular, according to subsumption and unification theories, explanations appeal to regularities that subsume what's being explained under some kind of law (e.g., Hempel & Oppenheim, 1948) or explanatory pattern (e.g., Friedman, 1974; Kitcher, 1989). In so doing, they relate the particular fact or observation to a generalization that supports further inferences (Lombrozo, 2006, 2012; Wellman & Liu, 2007). For example, by explaining Socrates' death by appeal to the consumption of a poisonous chemical contained within hemlock (i.e., coniine), one implicitly invokes the generalization that the chemical can cause death in humans. This generalization in turn supports predictions about the consequences of future coniine consumption, provides guidance about how to avoid a particular kind of death (i.e., don't consume hemlock), and even supports counterfactuals about how things could have been otherwise (e.g., if Socrates hadn't consumed hemlock, or if he'd had an antidote to coniine, he would have lived to see another day).

If explanations typically subsume what is being explained under some generalization, then engaging in explanation could influence learning and inference by driving reasoners to form broad generalizations and to consult them as a basis for further reasoning (Lombrozo, 2012). Consistent with this idea, research with adults has shown that prompts to explain can promote the discovery and extension of broad patterns that govern membership in novel categories (e.g., Williams & Lombrozo, 2010; Williams & Lombrozo, 2013; Williams, Lombrozo, & Rehder, 2013; see also Chi, DeLeeuw, Chiu, & LaVancher, 1994). Recent developmental work likewise suggests that when prompted to explain, even young children are more likely to favor broad patterns (Walker et al., 2012; Walker et al., submitted for publication) and to develop abstract theories, such as a theory of mind (Amsterlaw & Wellman, 2006), that can accommodate otherwise-puzzling observations (e.g., a character looking for an object in the wrong location). For example, Walker et al. (2012), Walker et al. (submitted for publication) found that when prompted to explain why particular types of objects activate a machine while others do not, preschool-aged children were more likely to rely on a feature that accounted for all observations (as opposed to a subset) in deciding which new objects were likely to activate the machine.

Many of the most far-reaching and useful generalizations are those that involve causal relationships, as they support interventions in addition to predictions. Generalizations relating hemlock and death (in the example with Socrates), or beliefs and behaviors (in theory of mind), are cases in point. Some accounts of explanation *require* that explanations be causal (e.g., Strevens, 2008; Woodward, 2005; Woodward, 2011), but one need not subscribe to a strictly causal theory of explanation to accommodate the observation that explanation and causation are often closely linked: the view that explanations privilege broad and useful generalizations is enough to support the idea that causation will often (if not always) be central to explanations. In line with this idea, previous research with adults has demonstrated that explanations help guide causal inferences (Heit & Rubinstein, 1994; Rehder, 2006; Sloman, 1994). There is also indirect evidence that causation is central to children's explanations (e.g., Hickling & Wellman, 2001). For example, young children's explanations often posit unobserved causes (Buchanan & Sobel, 2011; Legare, 2012; Legare, Gelman, & Wellman, 2010; Legare, Wellman, & Gelman, 2009), and Legare and Lombrozo (2014) found that children who explained learned a novel toy's causal (functional) mechanism (i.e., interlocking gears make a fan turn), but not other superficial properties (i.e., the color of the gears), more readily than children who did not. In the experiments that follow, we focus on causality as a canonical, inductively-rich property that's likely to be privileged in explanation, and we investigate the prediction that prompting young children to explain will help them appreciate and use causal similarities as a basis for learning and inference.

1.2. Inductive generalization: a shift from perceptual to conceptual?

A large body of research has examined the role of obvious (perceptual) properties versus non-obvious (hidden or abstract) properties, such as causal affordances, in guiding children's inductive inferences (e.g., Gelman, 2003; Gelman & Markman, 1986; Gelman & Markman, 1987; Gopnik & Sobel, 2000; Keil, 1989; Keil & Batterman, 1984; Nazzi & Gopnik, 2000; Newman, Herrmann, Wynn, & Keil, 2008). This research demonstrates that even young children are able to use both perceptual and non-perceptual properties in categorizing objects (e.g., Gelman & Markman, 1987; Gopnik & Sobel, 2000). Nonetheless, young children tend to spontaneously focus on highly salient surface features. Specifically, while older children and adults often group objects according to complex cues such as common internal properties, labels, and causal affordances, regardless of perceptual similarity (Carey, 1985; Keil, 1989; Medin, 1989; Rips, 1989), young children tend to group objects based on perceptual similarity, and only later shift to favoring other properties (e.g., Gelman & Davidson, 2013; Gentner, 2010; Keil & Batterman, 1984).

To illustrate, consider the findings from Nazzi and Gopnik (2000). In this study, children observed four objects placed on a toy, one at a time. Two of these objects were shown to be causally effective – they made the toy play music – and two were inert. One of the causal objects was then held up and labeled (e.g., "This is a Tib"), and children were asked to give the experimenter the other object with the same label (e.g., the other "Tib"). In conflict trials, the same perceptual properties appeared across causal and inert objects, and performance on such trials revealed a developmental shift: when generalizing the novel label, 3.5-year-olds relied on perceptual cues over causal cues, while 4.5-year-olds relied on causal cues over perceptual cues.

Between the ages of 3 and 5, children also shift how they generalize internal or hidden parts. For example, Sobel, Yoachim, Gopnik, Meltzoff, and Blumenthal (2007) used a procedure similar to that of Nazzi and Gopnik (2000) to demonstrate that older children (4-year-olds), but not younger children (3-year-olds), are more likely to infer that objects have shared internal parts when they share causal properties than when they share external appearance. These examples – and many others (e.g., see evidence from research on psychological essentialism: Gelman, 2003; Keil, 1989) – demonstrate that by 5 years of age, children begin to reliably favor inductively rich properties, such as common causal affordances, over perceptual similarity when generalizing from known to unknown cases.

There have been a variety of proposals for how best to characterize and explain this shift in children's inductive generalizations. For example, one possibility is that children first categorize objects by relying on perceptual or "characteristic" properties, and then shift to a different basis for categorizing objects, one based on more complex or "defining" properties (see Keil & Batterman, 1984). Another possibility is that the basic mechanism underlying children's judgments remains constant, but that the exercise of this mechanism results in different judgments as children gather new evidence. Specifically, properties are often encountered in correlated clusters, with perceptual information serving as a reliable indicator of other properties. As a result, perceptually-based judgments may be quite reasonable until sufficient evidence has been amassed to suggest an alternative (Gopnik & Sobel, 2000; Keil, 1989; Nazzi & Gopnik, 2000; Sobel et al., 2007). From this perspective, even very young children may already be equipped with the conceptual resources to reason on the basis of non-perceptual properties, including causal affordances, even though performance on various tasks can change in the course of development. Consistent with this idea, Gopnik and Sobel (2000) found that when presented with conflicting cues, younger children produced a variety of memory errors that indicated an assumed correlation between different types of properties (i.e., perceptual and causal), even when no such correlation existed in the data. Even looking-time data from infants suggests that by 14- to 18-months, children differentially attend to various perceptual and non-perceptual properties in different tasks (Booth & Waxman, 2002; Mandler & McDonough, 1996; Newman et al., 2008).

In four experiments, we examine the possibility that by 3 years of age, children *already have* the conceptual resources to generalize on the basis of inductively rich properties, and that their failure to do so often results from a failure to access or apply what they know. (For related arguments in other tasks and domains, see, e.g., Hood, Cole-Davies, & Dias, 2003; Kirkham, Cruess, & Diamond, 2003; Munakata, 2001; Sobel & Kirkham, 2006; Walker & Gopnik, 2014; Zelazo Zelazo, Frye, & Rapus, 1996.) We investigate whether the process of seeking or generating explanations facilitates access to and application of causal knowledge, supporting children's ability to reason on the basis of non-obvious but inductively rich causal properties as opposed to salient but superficial perceptual properties.

1.3. Overview of experiments

In the following experiments, we use a method similar to Nazzi and Gopnik (2000) and Sobel et al. (2007) to examine whether generating explanations makes children more likely to infer that an object's internal parts will be shared by other objects with common causal affordances as opposed to similar appearances (Experiments 1a and 1b), and more likely to believe that objects belong to the same category when they share common causal affordances as opposed to perceptual appearances (Experiment 2). In Experiment 3, we examine whether effects of explanation extend to lower-level cognitive processes, such as attention and memory, and whether they derive from a special relationship between explanation and inductively rich properties or from a global boost in performance. Together, these experiments provide insight into the role of explanation in causal inference in early childhood.

2. Experiment 1a

Experiment 1a examines whether explanation influences preschoolers' extension of a hidden, internal property to other objects that share either perceptual or causal properties. Children observed four sets of three objects that were individually placed on a toy that played music when "activated" (see Gopnik & Sobel, 2000). Each set contained three objects: one that activated the toy (*target object*), one that was perceptually identical to the *target object*, but failed to activate the toy (*perceptual match*), and one that was perceptually dissimilar to the *target object*, but successfully activated the toy (*causal match*). After each outcome was observed, children were asked to either explain (*explain* condition) or report (*control* condition) that outcome. Next, children received additional information about the target object: an internal part was revealed. Children were asked which one of the two other objects in the set (i.e., the *perceptual match* or *causal match*) shared the internal property with the *target object*. This method pit highly salient perceptual similarity against shared causal properties; children could base their generalizations on either one, but not both.

Given the hypothesis that generating explanations encourages learners to favor broad generalizations, and thus to focus on inductively-rich properties such as causal affordances, we predicted that children who were asked to explain each outcome would be more likely than children in the control condition to select the *causal match* over the *perceptual match*.

2.1. Methods

2.1.1. Participants

A total of 108 children were included in Experiment 1a, with 36 3-year-olds ($M = 40.9$ months; $SD = 3.7$, range: 35.8–47.7), 36 4-year-olds ($M = 53.3$ months; $SD = 3.6$, range: 48.5–59.8), and 36 5-year-olds ($M = 64.4$ months; $SD = 3.0$, range: 60.1–70.4). Eighteen children in each age group were randomly assigned to each of the two conditions (*explain* and *control*). There was no significant difference in age between the conditions, and there were approximately equal numbers of males and females assigned to each group. Five additional children were tested, but excluded due to failure to attend to the experimenter or complete the study. Children were recruited from urban preschools and museums, and a range of ethnicities resembling the diversity of the population was represented.

2.1.2. Materials

The toy was similar to the "blicket detectors" used in past research on causal reasoning (Gopnik & Sobel, 2000), and consisted of a 10″ × 6″ × 4″ opaque cardboard box containing a wireless doorbell that was not visible to the participant. When an object "activated" the toy, the doorbell played a melody. The toy was in fact surreptitiously activated by a remote control.

Twelve wooden blocks of various shapes and colors were used (see Fig. 1). A hole was drilled into the center of each block. Eight blocks contained a large red plastic map pin glued inside the hole; the remaining four blocks were empty. All of the holes were covered with a dowel cap, which covered the opening to conceal what was inside. Each of the four sets of blocks was composed of three individual blocks. Within each set, two blocks were identical in color and shape, and one of these (the *target object*) contained a map pin. The other block (the *perceptual match*) did not. The third block (the *causal match*) was perceptually dissimilar to the other two.

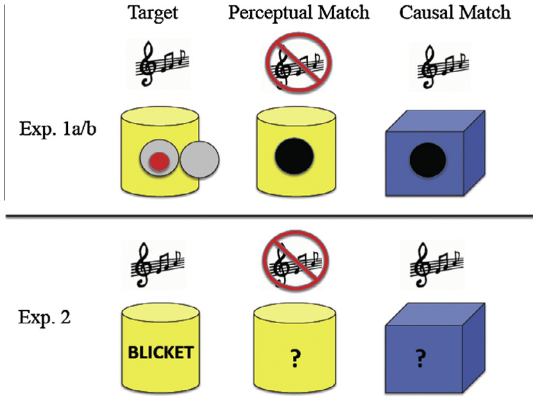

Fig. 1. Sample set of objects used in Experiments 1a/1b (top) and Experiment 2 (bottom). Each of the objects in Experiment 1a/b included a door (indicated by a black circle), which covered the internal part contained inside. Each row corresponds to a single set of items. There were a total of four sets of stimuli.

2.1.3. Procedure

Children participated in a brief warm-up game with the experimenter. Following this warm-up, the toy was placed on the table. The child was told, "This is my toy. Some things make my toy play music and some things do not make my toy play music." Then the first set of three blocks was brought out and placed in a row on the table. The order of presentation of the three blocks was randomized. One at a time, the experimenter placed a block on the toy. Two of the three blocks in each set (the *target object* and the *causal match*) caused the toy to activate and play music. The *perceptual match* did not. After children observed each outcome, they were asked for a verbal response. In the *explain* condition, children were asked to explain the outcome: "Why did/didn't this block make my toy play music?" In the *control* condition, children were asked to report the outcome (with a yes/no response): "What happened to my toy when I put this block on it? Did it play music?" After all three responses had been recorded, the experimenter demonstrated each of the three blocks on the toy a second time to facilitate recall.

Next the experimenter pointed to the set of objects and said, "Look! They have little doors. Let's open one up." The experimenter selected the *target object* and removed the cap to reveal the red map pin that had been hidden inside. The experimenter said, "Look! It has a little red thing inside of it. Can you point to the other one that also has something inside?" Children were then encouraged to point to one of the two remaining objects (i.e., the *perceptual match* or the *causal match*) to indicate which contained the same inside part, and this selection was recorded. Children could either select the block that was perceptually identical to the target *or* the object that shared the causal property, but not both.

Following their selection, children were not provided with feedback, nor were they allowed to explore the blocks. Instead, all blocks were removed from view, and the next set was produced. This procedure was repeated for the three remaining sets. Each child participated in a total of four trials, including a total of four unique sets of objects.

2.1.3.1. Coding. For each set of objects, children were given a score of "1" for selecting the *causal match* and a "0" for selecting the *perceptual match*. Each child could therefore receive between 0 and 4 points across the 4 trials. The explanations that children provided were also coded into five mutually-exclusive types: (1) appearance (e.g., *"It made the toy play music because it's purple," "...because it's round," "...because it looks like an apple"*), (2) internal parts (e.g., *"...because it has something inside of it," "...because it has a red thing in it," "...because it has batteries," "...because it has a motor"*), (3) kind (e.g., *"...because it's the right kind," "...because it's a music-maker," "...because it's musical"*), (4) other/non-informative (e.g., *"...because it's magic," "...because it wants/likes to," "...because it's special"*), and (5) no guess (e.g., *"I don't know"*). For the few participants who provided explanations that included both perceptual and internal properties, explanations were coded as appealing to internal properties. Because many of the children's explanations were quite minimal (only a couple of words in some cases), we did not examine the quality of children's responses beyond classifying them as belonging to particular explanation type.

Children's responses to the test questions were recorded by a second researcher during the testing session, and all sessions were video recorded for independent coding by a third researcher who was naïve to the hypotheses of the experiment. Interrater reliability was very high; the two coders agreed on 99% of the children's responses to the test questions and on 91.8% of children's explanations. Disagreements were resolved by a third party.

2.2. Results and discussion

Preliminary analyses revealed no trial-by-trial learning across the four sets of objects; children were no more likely to select the causal match on later trials than on earlier trials, Cochran's $Q(3) = 5.36$, $p = .148$. The data from the four trials were therefore combined to yield a single combined score that ranged from 0 to 4, and the data were analyzed with a 2 (condition) × 3 (age group) ANOVA (see Fig. 2). The ANOVA revealed main effects of condition, $F(1,102) = 50.70$, $p < .001$, and age, $F(2,102) = 7.34$, $p < .01$, with no significant interaction. Overall, children who were asked to explain ($M = 2.98$, $SD = 1.23$) were more likely than children in the control condition ($M = 1.61$, $SD = 1.58$) to generalize the internal part of the *target object* to the *causal match* as opposed to the *perceptual match*. To better understand the main effect of age, we conducted pairwise comparisons between age groups, which revealed no difference in performance between 3- and 4-year-olds, $p = .86$, but that 3- and 4-year-olds each selected the *causal match* significantly less often than 5-year-olds, $p < .01$.

We also conducted one-sample *t*-tests comparing performance to chance to assess whether explaining prompted children to override a preference to generalize on the basis of perceptual similarity. The 3-year-olds and 4-year-olds in the control condition selected the *perceptual match* significantly more often than chance, $t(17) = -3.69$, $p < .01$, and $t(17) = -2.53$, $p < .05$, respectively, while those in the explain condition selected the *causal match* significantly more often than chance, $t(17) = 3.01$, $p < .01$, and $t(17) = 2.48$, $p < .05$, respectively. Five-year-olds in the control condition performed no differently from chance ($M = 2.61$, $SD = 1.72$), $t(17) = 1.51$, $p = .15$, while 5-year-olds in the explain condition selected the *causal match* significantly more often than expected by chance ($M = 3.39$, $SD = 1.29$), $t(17) = 4.57$, $p < .001$.

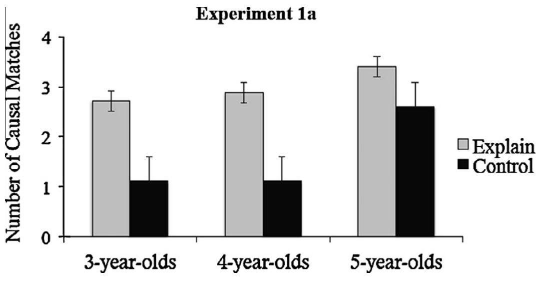

Fig. 2. Average responses in *explain* and *control* conditions for Experiment 1a. Higher numbers indicate a larger number of trials (of 4) on which an internal part was generalized in line with a shared causal property over perceptual similarity. Error bars correspond to one SEM in each direction.

These data suggest that in the absence of an explanation prompt, children relied primarily on the target object's salient perceptual features to predict whether a novel object would share an internal property. However, when children of the same age were asked to generate an explanation, they instead privileged the target object's causal efficacy in making inferences about internal properties.

2.2.1. Content of explanations

The frequencies with which children produced explanations of different types are reported in Table 1.

Baseline explanations for the first set of objects (before receiving any information about the internal properties) most often appealed to appearance (38%), with a minority (5%) appealing to internal properties. After observing the presence of the internal property for the first set of objects, explanations for the second set of objects appealed to appearance (33%) and internal properties (32%) equally often. By the final set, explanations most often appealed to internal parts (38%). An exact McNemar's test comparing the proportion of explanations that appealed to internal parts across the first and last trials revealed a significant difference, $p < .0001$.

Although we did not code the "quality" of children's explanations, we did examine the relationship between explanation type and performance. To do so, we identified the type of explanation that each child produced most often (i.e., the modal explanation for each child; see Table 2) and analyzed generalizations as a function of this designation. Overall, children who provided internal explanations as their modal response – arguably the most

Table 1
Frequency of explanation types for each set in Experiments 1a, 1b, and 2.

	Set 1	Set 2	Set 3	Set 4	Total
Exp. 1a					
Appearance	61	53	38	41	193
Internal	8	51	55	61	175
Kind	8	4	10	11	33
Other	32	23	26	16	97
No guess	53	31	33	33	150
Exp. 1b					
Appearance	17	14	–	–	31
Internal	2	17	–	–	18
Kind	3	4	–	–	7
Other	16	9	–	–	25
No guess	16	10	–	–	26
Exp. 2					
Appearance	50	41	45	44	180
Internal	36	15	24	12	87
Kind	13	15	12	11	51
Label	0	9	14	17	40
Other	26	44	40	52	162
No guess	38	39	28	23	128

Table 2
Proportion of causal matches in Experiments 1a and 2 as a function of child's modal explanation type.

Modal explanation	Frequency	% Causal matches
Exp. 1a		
Appearance	13	53
Internal	24	80
Kind	9	63
Other	4	45
No guess	2	88
No mode	2	88
Exp. 2		
Appearance	16	33
Internal	6	33
Kind	2	88
Label	4	100
Other	14	48
No guess	11	48
No mode	1	75

Note: The number of children designated in each category is reported under "frequency."

relevant explanation in this task – were significantly more likely to select causal matches than the aggregate of other children (80% versus 60%), $t(79) = 1.99$, $p = .05$. Despite the limitations associated with combining all other explanation types in a single group (which was necessary due to the small sample sizes), these results suggest that children who provided the most relevant explanation may have benefited most from the explanation prompt. We also found that 4- and 5-year-olds were each more likely to provide modal explanations that appealed to internal parts (20% and 19% of explanations, respectively) than 3-year-olds (6% of explanations), $\chi^2(54,1) = 5.25$, $p < .05$ and $\chi^2(54,1) = 4.29$, $p < .05$, respectively.

We also found evidence that the prompt to explain impacted children's inferences even when the explanations that were generated did not appeal to internal properties. For example, the two children who provided no modal explanation (i.e., children who provided distinct explanation types for each set) and the two children who provided a modal explanation of "no guess" were (numerically) the most likely to select the causal match (88% each). In fact, each category of modal explanation, regardless of type (appearance: 53%, kind: 63%, other: 45%), was associated with a higher proportion of *causal matches* than that observed of children in the control condition (40%). Combining all of the children who provided modal explanations other than insides into a single group and comparing their responses to those of children in the control condition revealed a significant difference, $t(54) = -2.19$, $p < .05$. These data suggest that although children who provided the "correct" (internal) explanation were more likely to generalize according to causal as opposed to perceptual similarity, simply receiving an explanation prompt was enough to impact children's reasoning in this task.

In sum, our data support the proposal that prompts to explain increase children's reliance on inductively rich properties (as opposed to merely perceptual ones) as a basis for inference, and further suggest that effects of explanation are not restricted to children who happen upon the "correct" explanation for the task. There is an alternative explanation for our findings, however, that should be addressed. It is possible that explanation promoted greater projection to the *causal match* because the experimenter revealed the internal property immediately after children were prompted for an explanation, encouraging them to interpret the reveal as the experimenter's means of providing an answer to the 'why" question the child had attempted to answer. Thus, the design of the task may have signaled to children that the internal part was the reason why the blocks made the toy play music (even if this information was not then reflected in all children's explicit explanations). In Experiment 1b we therefore investigate whether children generalized the internal property to the causal object because the timing and context of the explanation prompt supported a particular pragmatic inference, or because the process of explaining itself directed children to posit or privilege causality as a basis for generalization.

3. Experiment 1b

The purpose of Experiment 1b was to rule out a pragmatic account of the findings from Experiment 1a. The procedure in Experiment 1b involved a critical modification from Experiment 1a: the addition of a second experimenter. Rather than having the same experimenter request explanations and reveal the internal properties of the objects, one researcher (R1) demonstrated the causal properties of the objects and provided the explanation prompt, and a second researcher (R2) (who had not observed the previous demonstration or explanation) revealed the internal part and solicited the generalization judgment.

If children in Experiment 1a who were prompted to explain preferentially generalized on the basis of causal properties because they took the researcher's revelation of the internal property as a potential answer to that researcher's why-questions, then changing researchers in this way should block the relevant pragmatic inference, and lead to performance comparable to the control condition. In contrast, if something about the process of explaining prompts children to privilege causal similarity over perceptual appearance in our task, then this change in task pragmatics should not change children's generalization judgments.

Because we planned to compare children's performance in Experiment 1b to performance in Experiment 1a, and because we found no age differences between 3- and 4-year-old children, we only included one subgroup of children: 4-year-olds in the *explain* condition. By comparing the performance of this new group of children with that of 4-year-olds in the *explain* and *control* conditions from Experiment 1a, we can assess whether the results of Experiment 1a were plausibly an artifact of the pragmatics of the task.

3.1. Methods

3.1.1. Participants

Eighteen 4-year-olds were included in Experiment 1b ($M = 53.14$ months; $SD = 3.1$, range: 48.8–59.4). All children were assigned to the *explain* condition. There was no significant difference in age between the 4-year-old children included in Experiments 1a and 1b, $p = .84$, and there were approximately equal numbers of males and females. Two additional children were tested, but excluded due to experimenter error. Recruitment procedures and demographics were identical to Experiment 1a.

3.1.2. Materials

Materials were identical to those used in Experiment 1a.

3.1.3. Procedure

The procedure was similar to the one used in the *explain* condition in Experiment 1a, with two exceptions. First, one researcher (R1) provided explanation prompts, while a different researcher (R2) revealed the hidden properties and solicited the generalization judgments. Second, there were only two trials (rather than four) to avoid the concern that repeatedly switching experimenters could make the experimental situation too implausible or complex.

After children observed the first set of three objects placed on the toy and provided explanations for each one to R1, R2 entered the testing room. R2 said, “Hey, cool! Can I look at those?” R1 consented and walked away from the table. R2 examined the blocks on the table, saying, “Look! They have little doors. Let’s open one up.” R2 then selected the *target object* and removed the cap to reveal the red map pin that had been hidden inside, saying, “Look! It has a little red thing inside of it. Can you point to the other one that you think also has something inside?” As in Experiment 1a, children were encouraged to point to one of the two remaining objects (i.e., the *perceptual match* or the *causal match*) to indicate which contained the same inside part, and this selection was recorded. Following their selection, children were not provided with feedback, nor were they allowed to explore the blocks. Instead, R1 returned to the table, R2 departed from the testing room, and all blocks were removed from view. This two-experimenter procedure was repeated for one additional set of blocks.

3.1.3.1. Coding. For each set of objects, children were given a score of “1” for selecting the *causal match* and a “0” for selecting the *perceptual match*. Each child could therefore receive between 0 and 2 points across the two trials. Explanation coding procedures were identical to Experiment 1a. Two coders agreed on all of the children’s responses to the test questions and on 94.4% of children’s explanations; disagreements were resolved by a third party.

3.2. Results and discussion

As in Experiment 1a, children in Experiment 1b did not perform significantly differently across trials, Cochran’s $Q(1) = .143$, $p = .705$. Data from both trials were therefore combined into a single score from 0 to 2, and the scores from this group were compared with the combined score from the first two trials of the 4-year-old participants in the *explain* and *control* conditions from Experiment 1a.

A univariate analysis of variance (ANOVA) with combined score as the dependent variable and condition (3: Exp. 1a *control*, Exp. 1a *explain*, Exp. 1b *explain*) as the independent variable revealed a main effect of condition, $F(2,54) = 7.79$, $p < .01$. Children who were asked to explain in both Experiments 1a ($M = 1.3$, $SD = .77$) and 1b ($M = 1.56$, $SD = .62$) were each more likely than controls ($M = .61$, $SD = .85$) to generalize the internal part of the *target object* to the *causal match* as opposed to the *perceptual match*, $p < .01$ and $p < .001$, respectively. Pairwise comparisons revealed no difference in performance between 4-year-olds in the *explain* conditions of Experiments 1a and 1b, $p = .379$. We also conducted a one-sample t-test comparing children’s performance to chance. Children in Experiment 1b selected the *causal match* significantly more often than chance, $t(17) = 3.83$, $p < .01$.

These data suggest that children in Experiment 1a were *not* simply interpreting the experimenter’s revelation of the internal property as an answer to the “why?” question that the experimenter had previously posed. In Experiment 1b, the experimenter who provided the explanation prompt was different from the experimenter who revealed the hidden property, so the relevant pragmatic inference was disrupted. Instead, it appears that children in the *explain* condition privileged the target object’s causal efficacy in making inferences about internal properties as a consequence of something about the very process of explaining.

3.2.1. Content of explanations

Frequency data for each explanation type are reported in Table 1. Explanations were divided into the same five categories as in Experiment 1a.

Baseline explanations for the first set of objects (before receiving any information about the internal properties) most often appealed to appearance (32%), with a minority (4%) appealing to internal properties. After observing the presence of the internal property, explanations for the second set of objects most often appealed to internal properties (32%), with explanations appealing to appearance dropping slightly (30%). An exact McNemar’s test comparing the proportion of explanations that appealed to internal parts across the first and second (last) trials revealed a significant effect, $p < .0001$. Because there were only two trials, an analysis of children’s modal explanation was not conducted.

4. Experiment 2

The purpose of Experiment 2 was twofold. First, we were interested in whether the effect of explanation on children’s inferences is restricted to generalizations concerning the relationship between causal properties and internal (or hidden) parts, or whether it extends to other properties as well. Second, we were specifically interested in whether explanation would affect how children extend novel labels. An effect of explanation on label extension would suggest that the process of explaining changes how children form categories, potentially shifting them from categories formed on the basis of perceptual properties to those tracking non-obvious, inductively rich causal properties. The ability to override perceptual similarity is an important hallmark of both scientific and everyday categories, as highly salient perceptual properties can be good predictors of category membership, but they can also be deceptive. For example, a dolphin may resemble a large fish, but dolphins are actually warm-blooded mammals. When such properties appear in conflict with one another, category membership is often based on non-obvious cues (e.g., internal biological properties) rather than surface appearance (e.g., having a tail).

Previous research demonstrates the importance of labels as indicators of category membership and guides to inference (e.g., Carey, 1985; Diesendruck, Markson, & Bloom, 2003; Gelman, 2003; Gelman & Markman, 1987; Gelman & Medin, 1993; Keil, 1989; Legare et al., 2010; Medin, 1989; Nazzi & Gopnik, 2000; Rips, 1989). For example, Gelman and Coley (1990) found that in some cases, even 2-year-old children answered questions in line with category membership over appearances when labels were provided. But in the absence of labels, judgments are typically dominated by perceptual similarity. In fact, some have argued that children's categories are driven by low-level perceptual mechanisms that lead them to focus on object shape and other surface features (e.g., Landau, Smith, & Jones, 1988). However, other findings suggest that children extend labels differently depending on their intuitions about the kinds of object being classified, or on the nature of the task, and that classification is not always perceptually driven (e.g., Carey, 1985; Diesendruck et al., 2003; Keil, 1989). Finding that a prompt to explain leads children to extend labels on the basis of common causal properties would further suggest that even young children are able to form categories that disregard appearances, and that explaining helps them do so.

In sum, Experiment 2 used a method similar to Experiment 1a to examine whether the effects of explanation would extend to children's generalization of a novel *label* from a target object to an object that was either perceptually similar or causally similar. We predicted that explaining would make children more likely to attend to the causal powers of objects, which in turn would make it more likely for children to use causal properties as a basis for extending category labels to novel objects.

4.1. Method

4.1.1. Participants

A total of 108 children were included in Experiment 2, with 36 3-year-olds ($M = 42.1$ months; $SD = 3.8$, range: 35.9–48.0), 36 4-year-olds ($M = 54.0$ months; $SD = 3.0$, range: 48.4–59.9), and 36 5-year-olds ($M = 65.0$ months; $SD = 3.8$, range: 60.6–70.9). Eighteen children in each age group were randomly assigned to each of the two conditions (*explain* and *control*). There were no significant differences in age between the conditions, and there were approximately equal numbers of males and females in each. Eight additional children were tested, but excluded due to failure to complete the study or failure to respond to the experimenter. Two more children were excluded due to experimenter error. Children were recruited from urban preschools and museums, and a range of ethnicities resembling the diversity of the population was represented.

4.1.2. Materials

The toy from Experiment 1 was again used in Experiment 2. Twelve wooden blocks of various shapes and colors were also used. There were a total of four sets of objects, each containing three blocks. As in Experiment 1, two of these blocks (the *target object* and the *perceptual match*) were perceptually identical (same color and shape) and one of these blocks (the *causal match*) was distinct (see Fig. 1).

4.1.3. Procedure

The procedure for Experiment 2 was identical to Experiment 1, with one exception: Instead of revealing a hidden internal property, the experimenter held up the *target object* and labeled it, saying, "See this one? This one is a blicket! Can you point to the other one that is also a blicket?"

4.1.3.1. Coding. Coding for Study 2 was identical to Study 1, with children receiving a "0" for generalizations to the perceptual match and a "1" for generalizations to the causal match, resulting in a score of 0–4 points across the four sets. Interrater reliability was very high; the two coders agreed on >99% of the children's responses to the test questions and 96.8% of children's explanations The few minor discrepancies were resolved by a third party.

4.2. Results and discussion

Preliminary analysis revealed no significant differences across trials, Cochran's $Q(3) = .60$, $p = .896$. Data from all four trials were therefore combined into a single score from 0 to 4 and analyzed with a 2 (condition) × 3 (age group) ANOVA (see Fig. 3). This analysis revealed a main effect of condition, $F(1,102) = 13.51$, $p < .001$, and no additional significant effects. Overall, children who were asked to explain ($M = 1.91$, $SD = 1.83$) were more likely than children in the control condition ($M = .72$, $SD = 1.47$) to generalize the label to the *causal match* as opposed to the *perceptual match*, regardless of age.

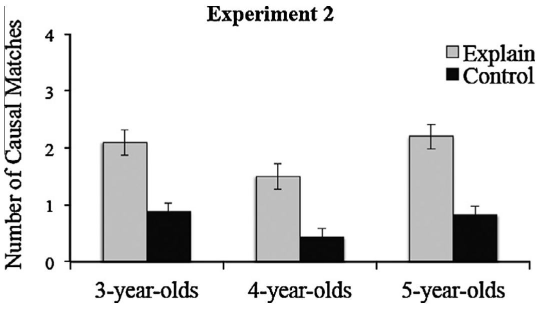

Fig. 3. Average responses in *explain* and *control* conditions for Experiment 2. Higher numbers indicate a larger number of trials (of 4) on which a label was generalized in line with a shared causal property over perceptual similarity. Error bars correspond to one SEM in each direction.

We next considered the data against chance responding. One-sample t-tests revealed that 3-, 4-, and 5-year-olds in the control condition selected the *perceptual match* significantly more often than chance, $t(17) = -2.93$, $p < .01$, $t(17) = -3.69$, $p < .01$, and $t(17) = -3.10$, $p < .01$, respectively. In the explanation condition, the average of children's selections did not differ significantly from chance, $t(17) = .12$, $p = .90$, $t(17) = -1.26$, $p = .23$, and $t(17) = .375$, $p = .712$, respectively. However, examining the distribution of selections across the four trials revealed that approximately half of the children in the explanation condition selected the causal match on three or more trials (50% for 3-year-olds, 44% for 4-year-olds, and 56% for 5-year-olds). This distribution differed significantly from that expected by chance in all age groups, $\chi^2(4) = 84.26$, $p < .001$, $\chi^2(4) = 66.49$, $p < .001$, and $\chi^2(4) = 83.97$, $p < .001$, respectively.

Because responses in Experiment 2 were not normally distributed, we conducted a non-parametric test comparing the performance of children across conditions. Children who selected the causal match on three or four trials were designated as "causal reasoners," and all others as "perceptual reasoners" (see Sobel et al., 2007). Results of a Chi-square test replicate the findings reported in the parametric tests above, revealing a significant effect of condition, $\chi^2(1) = 8.28$, $p < .01$, with children in the explanation condition more likely to be designated "causal reasoners" (50%) than children in the control condition (19%).

In sum, like the younger children in Experiment 1a, children in the control condition in Experiment 2 relied primarily on a target object's salient perceptual features to predict whether a novel object would share a category label. This is particularly surprising given that the same label was provided across all four trials, during which the perceptual features of the target object varied from trial to trial. However, when children of all ages were asked to generate an explanation for the evidence that they observed, they considered the target object's causal efficacy significantly more often in making inferences about shared labels.

4.2.1. Content of explanations

Explanations were coded as in Experiments 1a and 1b, with one additional explanation type for those children who appealed to the label (e.g., *"It's a blicket"*) (see Table 1). Appearance explanations were most common overall (28% of all explanations); however, there was an increase in explanations that explicitly mentioned the label across trials, with 0% in the first set and 11% in the final set. An exact McNemar's test comparing the proportion of label explanations across the first and last sets revealed a significant difference, $p < .0001$.

To analyze the relationship between explanation type and performance in Experiment 2, we again calculated a modal explanation for each child, reflecting the most common explanation type that the child provided (see Table 2). Children who most often provided an explanation that referred to the label also privileged causality in generalizing the label more often (100% versus 39%). However, so few children appealed to labels as their modal explanation ($N = 4$) that there were no significant differences in performance as a function of modal explanation type.

Also as found in Experiment 1a, simply being prompted for an explanation was enough to affect children's inferences. Each modal explanation, regardless of type (appearance: 33%; internal: 33%; kind: 88%; other: 48%; no guess: 48%), was associated with a greater probability of selecting the *causal match* than in the control condition (18%). Combining all of the children who provided modal explanations other than labels into a single group and comparing their responses to those of children in the control condition revealed a significant difference, $t(90) = 2.39$, $p < .02$. As in Experiment 1a, these data suggest that although providing the most relevant explanation type (in this case, an appeal to the category label) leads to a special boost in performance, simply receiving an explanation prompt is enough to influence reasoning.

4.2.2. Comparing Experiments 1 and 2

To examine differences across our two experiments, we analyzed the data from Experiments 1a and 2 in an ANOVA with experiment as a between-subjects factor. This analysis revealed a significant difference in children's performance in Experiments 1a and 2, with a greater number of causal responses in Experiment 1 ($M = 2.3$; $SD = 1.6$) than Experiment 2 ($M = 1.31$; $SD = 1.8$), $F(1) = 22.41$, $p < .001$. There were also significant effects of age, $F(2) = 4.74$, $=p < .02$, and condition, $F(1) = 38.0$, $p < .001$, but no significant interactions. In other words, despite a greater baseline tendency to privilege perceptual features when reasoning about labels than about insides, the effect of explanation – increasing causal responding – did not differ across our two experiments, nor across age groups.

The observed difference in children's baseline responding across our two experiments is in line with previous research (Gopnik & Sobel, 2000; Sobel et al., 2007), which has found that children are more willing to privilege causality over appearances when extending internal parts than when extending labels. This pattern could also reflect a tension between more conceptual uses of labels, such as reference to essences or causes, and the more perceptually-based "shape-bias" found in noun labeling (e.g., Gelman, 2003; Gelman & Markman, 1986; Imai, Gentner, & Uchida, 1994; Jones & Smith, 1993; Landau et al., 1988; Smith, Jones, & Landau, 1996). Nevertheless, explanation has a similar effect in promoting more conceptual (as oppose to perceptual) generalizations for both insides and labels. In effect, children in Experiment 2 were categorizing differently, depending on whether they explained or not. These results show that explanation guides children to attend to causal properties as an important but non-obvious basis for category membership.

5. Experiment 3

The findings from Experiments 1a, 1b, and 2 confirm our prediction that explanation encourages children to favor inductively rich properties (i.e., causality) as a basis for generalization. In Experiment 3 we hoped to bolster and further develop our interpretation of these novel findings by investigating three specific questions. First, the previous experiments demonstrate that explanation encourages children to privilege causal properties over perceptual properties when it comes to generalizing insides or labels. We propose that this is because the process of generating explanations prompts learners to seek broad, generalizable patterns, and that this in turn should privilege properties that feature in such generalizations – by definition, those that are inductively rich. In Experiment 3, we investigate whether effects of explanation are restricted to inductive generalizations, or additionally manifest in lower-level processes that might be prerequisites to inductive inference, such as memory for object properties. In particular, might prompting children to explain make them more likely to attend to, and therefore effectively *remember*, an object's causal properties? And will benefits for memory be restricted to causal properties, which is directly related to what's being explained (i.e., an effect or its absence), or will they extend to other inductively rich properties that might figure in the explanations themselves, such as insides and category label?

Second, if we do find that explanation improves memory for properties such as insides and labels, it raises a question about the selectivity of explanation's effects (see also Legare & Lombrozo, 2014). In particular, the findings from the preceding experiments are consistent with the idea that prompts to explain result in an indiscriminate increase in children's overall attention or engagement, which could potentially account for more adult-like performance without needing to posit a special relationship between explanation and inductively rich properties. This account, like ours, would predict that children who are prompted to explain would have better memory for object insides and labels than those in a control condition, but would additionally predict that children who explain should have better memory for a property that is *not* inductively rich. In Experiment 3, we introduce such a property in the form of a sticker that is not correlated with any other object properties. Our hypothesis suggests that effects of explanation are selective – as opposed to indiscriminate – and predicts improved memory for object insides and labels (which are correlated with causal properties in both the task and in the world), but not for an uncorrelated perceptual property like the sticker.

A final question addressed by Experiment 3 is whether explanation-induced advantages for inductively rich properties come at the *expense* of memory for other kinds of properties. In particular, it could be that explainers simply fail to remember an uncorrelated sticker any better than controls, or that they actually show *impairment* in memory for this feature relative to controls. The latter possibility is consistent with previous research involving both children (e.g., Legare & Lombrozo, 2014) and adults (e.g., Hegarty, Mayer, & Monk, 1995; Needham & Begg, 1991) in which increased focus on an important abstract principle decreases memory for surface features.

To test these ideas, children in Experiment 3 were asked to explain or report causal outcomes after observing four unique objects, two of which activated the toy. After each object was placed on the toy, three properties were revealed: an internal part, a label, and a sticker (added to the object). The internal parts and the labels correlated with the toy's activation (i.e., all and only objects that activated the toy had a particular inside part and label) while the sticker did not. Children then completed a memory task in which they were asked to report the properties of each object. Because we did not observe age differences in the effects of explanation in Experiments 1–2, Experiment 3 was restricted to 4-year olds.

5.1. Method

5.1.1. Participants

A total of 36 4-year-olds were included in Study 3 ($M = 53.8$ months; $SD = 3.7$ months; range = 47.9–59.7). Eighteen children were randomly assigned to each of two conditions (*explain* and *control*). There were no significant differences in age between the conditions, and there were approximately equal numbers of males and females in each. Three additional children were tested but excluded due to experimenter error. Children were recruited from urban preschools and museums, and a range of ethnicities representative of the diversity of the population participated.

5.1.2. Materials

Experiment 3 used the same toy as in the previous experiments. A different set of test blocks was used, however, which consisted of 4 *unique* blocks – i.e., each block was distinct in color and in shape (see Table 3). As in Experiments 1a and 1b, all blocks had a hole drilled into the center. Two of the blocks had a red, round plastic map pin glued inside and two of the blocks had a white, square eraser glued inside the hole. Four stickers were used during the experiment (two heart stickers and two star stickers). Several small cards were constructed as memory aids for use during the test phase of the experiment. Half of the cards had an image of a black music note (placed in front of the objects that children believed activated the toy), and half of the cards had an image of a black music note crossed out with a red "X" (placed in front of the objects that children believed did not activate the toy). Four additional cards were constructed: one with a red circle, one with a white square, one with a heart sticker, and one with a star sticker. These cards were used to facilitate the forced-choice test.

5.1.3. Procedure

As in the previous experiments, the experimenter introduced the toy. The experimenter then produced a single block and placed it on the toy. The child observed as the block did or did not cause the toy to play music. As before, children in the *explain* condition were asked to explain the outcome for each of the blocks and children in the *control*

Table 3
List of properties for objects used in Experiment 3.

	Object 1	Object 2	Object 3	Object 4
Causal	Yes	No	Yes	No
Internal	Red	White	Red	White
Label	"Toma"	"Fep"	"Toma"	"Fep"
Sticker	Heart	Heart	Star	Star

condition were asked to report the outcome with a "yes/no" response. After the response was recorded, the experimenter repeated the demonstration a second time.

The experimenter then provided three additional pieces of information about the object: the type of internal part was revealed ("Look! It has a little door on it! Let's open it up. Look, there is a [red]/[white] thing inside."), a label was provided ("See this one? This one here? This one is a [Fep]/[Toma]!"), and a sticker was placed on the bottom ("Now I am going to put a sticker on it! I am going to put a [heart]/[star] sticker on the bottom, see?"). The experimenter repeated each property twice, and then the block was removed from view. The entire procedure was repeated for the three remaining blocks, one at a time. All children observed the causal property first. The order of the remaining three properties was counterbalanced.

Next, the experimenter placed all four objects on the table in front of the child in random order, and told the child that they would now play a "memory game." Children were asked a baseline causal memory question first, and then three additional property memory questions in randomized order. To assess baseline recall for the causal property of each object, the experimenter produced two cards – one with a music note, and one with a crossed out music note. The experimenter asked the child to point to the card that indicated whether the block did or did not play music. The child responded once for each of the four objects. Depending upon the child's response, the experimenter would then place an additional card (with a music note or a crossed-out music note) in front of the object, which would remain throughout.

To assess recall for the internal part, the experimenter produced two cards – one with a red circle and one with a white square. The experimenter asked the child to point to the card that indicated the type of thing inside the block. The child responded once for each of the four objects. To assess recall for the label, the experiment said, "Some of these blocks were called 'Tomas' and some of these blocks were called 'Feps'. What was this one called, a 'Toma' or a 'Fep'?" The child responded once for each object. The order of presentation was counterbalanced across trials.

Finally, to assess recall for the type of sticker added to the block, the experimenter produced two cards – one with a heart sticker and one with a star sticker. The experimenter asked the child to point to the card that indicated the type of sticker added to the bottom of the block. The child responded once for each of the four objects.

5.1.3.1. Coding. Memory for internal parts, labels, and stickers was solicited in the same order as the corresponding properties were presented to that child in the demonstration phase of the experiment. For each property, children were given a score of "1" for accurate recall and a "0" for inaccurate recall. Because there were a total of four objects, each child could receive between 0 and 4 points for each property.

5.2. Results and discussion

Because the causal property was always presented first during the observation phase, and always assessed first during the testing phase, memory for the objects' causal properties was analyzed separately with a one-way ANOVA. Results of this ANOVA revealed that while the majority of children in both conditions were able to recall the causal property of each object, children in the explain condition were significantly more accurate ($M = 3.93$, $SD = .24$) than controls ($M = 3.39$, $SD = .78$), $F(1,34) = 8.42$, $p < .01$.

A repeated measures ANOVA with the other object properties (internal part, label, sticker) as the repeated measure and condition (explain, control) and order of presentation (label-sticker-insides, insides-label-sticker, sticker-insides-label) as the between subjects variables revealed a main effect of object property, $F(2,60) = 7.05$, $p < .01$, as well as the predicted interaction between object property and condition, $F(2,60) = 8.23$, $p < .002$ (see Fig. 4). Children who were prompted to explain were significantly more accurate than controls in reporting the labels, $F(1,34) = 9.34$, $p < .01$, but *less* accurate than controls in recalling the sticker type, $F(1,34) = 5.16$, $p < .05$. Although children who explained were numerically more accurate in recalling the internal part than controls ($M = 3.06$, $SD = 1.3$ and $M = 2.78$, $SD = 1.0$, respectively), this difference was not significant, $F(1) = .536$, $p = .47$.

These data address all three of the questions raised in Experiment 3. First, explanation does have an influence on memory for different object properties, and is not limited to inductive generalizations. Second, the findings challenge the idea that engaging in explanation simply improves overall engagement or attention in an indiscriminate manner. Instead, these data support the proposal that children who explain are more likely to *selectively* recall inductively rich, correlated cluster of properties (causality, internal part, label). Finally, we also found that children who were prompted to explain were significantly *less* likely to recall a superficial, perceptual property that did not correlate with other features, suggesting that benefits of explanation can come at a cost.

It is noteworthy that children who explained could not have been simply *ignoring* superficial perceptual features altogether, since the only way to track which properties corresponded to which objects in our task was to recall the unique color and shape of each of the blocks. Instead, explanation appears to impair memory for uncorrelated properties – those that are unrelated to other properties and therefore unlikely to support generalizations.

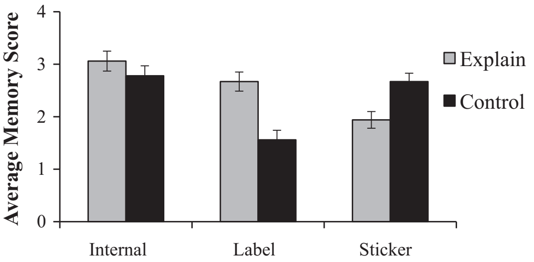

Fig. 4. Average memory score (out of 4 trials) for each property assessed in Experiment 3. Error bars correspond to one SEM in each direction.

5.2.1. Content of explanations

Children's explanations were coded according to the categories generated in Experiment 2, with the addition of a new possible category: appeal to the sticker. Combining explanation data from all four objects (a total of 76 individual explanations), there were a total of 24 explanations that appealed to appearance, 27 explanations that appealed to the internal part, 4 explanations that appealed to the kind of object, and 2 explanations that appealed to the label. Notably, however, *none* of the children's explanations appealed to the presence of the sticker. This provides additional support for the claim that explanation selectively increased attention to those properties that were inductively rich. An exact McNemar's test comparing the proportion of internal property explanations across the first (0%) and last sets (53%) revealed a marginally significant difference, $p = .06$ (one-tailed).

6. General discussion

Our data demonstrate that prompting young children to explain makes them more likely to privilege inductively rich, non-obvious causal properties over salient surface similarity in making novel inferences. Children in the *control* conditions, who were not prompted to explain, instead based their judgments on perceptual similarity. These effects of explanation cannot be explained by the pragmatics of the task, as explanation produced the same effect when a two-experimenter design was employed (Experiment 1b). Moreover, these effects of explanation were not restricted to a particular kind of inference, as comparable effects were observed across two quite different judgments: the generalization of hidden, internal parts (Experiment 1a) and inferences about category membership (Experiment 2). Finally, these effects were not restricted to a particular age group: we found comparable effects of explanation across our 3-, 4-, and 5-year-old participants.

The results of Experiment 3 provide additional support for the idea that explanation privileges inductively rich properties, demonstrating improved memory for a correlated cluster of such properties (not just for causal affordances) in children prompted to explain. Importantly, Experiment 3 also provides evidence that effects of explanation are selective: Children who explained had *impaired* memory for an uncorrelated superficial property (the sticker). This challenges one possible alternative interpretation of the results: that explanation produces a general benefit for learning by globally and indiscriminately increasing engagement or motivation (see also, Legare, 2012; Legare & Lombrozo, 2014), and additionally suggests that the benefits of explanation are not without costs (see also Williams et al., 2013).

The present findings suggest that children as young as 3 years of age have the conceptual resources to reason on the basis of non-obvious properties, such as causal affordances. These findings are therefore consistent with others suggesting children's early competence (e.g., Booth & Waxman, 2002; Gopnik & Sobel, 2000; Mandler & McDonough, 1996; Newman et al., 2008). Nonetheless, young children tend to privilege perceptual features over these less obvious alternatives under most conditions (e.g., Gelman, 2003; Keil, 1989; Sobel et al., 2007; Wellman & Gelman, 1992), and our findings go beyond prior work to identify a novel process that helps children overcome this tendency: namely engaging in explanation. In other words, engaging in explanation appears to facilitate children's access to (or ability to use) knowledge concerning the inductive relevance of causal properties.

Our findings have additional potential implications for our understanding of conceptual development. Experiments 1 and 2 deliberately spanned the age range (3–5 years) over which prior studies – which involved no explanation prompts – found developmental changes in children's tendency to generalize on the basis of perceptual versus causal properties (Nazzi & Gopnik, 2000; Sobel et al., 2007). While we did find age-related changes in children's baseline tendency to generalize one way or the other, the effects of explanation were uniform across ages. That is, we did not find interactions between the explanation manipulation and age group. One possibility is that the differences *within* age-groups observed across our experimental groups were driven by distinct mechanisms from those governing the changes observed *across* age-groups in our study and in others. For example, while the experimental effects were driven by explanation, the developmental effects could have been driven by general improvements in executive function or inhibitory control, or by different intuitive theories at different points in development. Another possibility, however, is that older children were more likely than younger children to engage in explanation *spontaneously* (i.e., in the absence of a prompt), shifting performance towards causal inferences in the control condition, and to generate more effective explanations when prompted to explain, leading to a comparable shift in the explanation condition. Consistent with these ideas, Legare and Lombrozo (2014) found that older children were more likely than younger children to generate explanations in response to an ambiguous verbal prompt, suggesting that self-initiated explanation increases over this age range. And in Experiment 1, we found that older children were more likely than younger children to provide explanations that appealed to internal parts, suggesting an age-related boost in explanation quality. Age-related changes in explanation frequency and quality could be driving part of the developmental shift in children's baseline tendency to generalize according to perceptual versus non-perceptual properties. The current data cannot adjudicate between these possibilities, but do raise them as promising hypotheses for future research.

We have discussed effects of explanation in Experiments 1 and 2 as favoring causal similarity over perceptual similarity, however, it is worth returning to the ideas about explanation that motivated our initial predictions, as they suggest a more nuanced view. We propose that explanations tend to subsume what is being explained under a pattern or regularity, and that in so doing, the act of explaining could lead children to recognize or formulate broad generalizations that in turn support inference to new cases (Legare, 2014; Lombrozo, 2012; see also Walker et al., 2012; Walker et al., submitted for publication; Wellman & Liu, 2007; Williams & Lombrozo, 2010; Williams & Lombrozo, 2013; Williams et al., 2013). On this view, explanation drives learners towards broad generalizations, not towards causal properties (or away from perceptual properties), per se. However, children may already have formed higher-level generalizations (Dewar & Xu, 2010; Kemp, Perfors, & Tenenbaum, 2007) suggesting that certain types of properties, such as insides and category labels, are more likely to track common causal properties than superficial perceptual ones.

Consistent with this idea, some existing findings support the proposal that internal properties have a special status relative to a superficial perceptual property, such as a sticker, even when their correlational structure is matched within the context of a specific task. Beyond our own findings from Experiments 1a and 2, Sobel et al. (2007, Experiment 3) report an experiment in which the researcher presented a target object that produced an effect and revealed *two* properties of the object: an internal part and a sticker affixed to its back. Four-year-olds, but not 3-year-olds, inferred that another object with the same internal part was more likely to produce the effect than an object that shared the same sticker. In other words, older children spontaneously favored an internal property over a temporary perceptual one as a basis for generalizing a causal property, even in the absence of explicit evidence that the internal property was more likely to be correlated with causality in the context of the experimental task. This suggests that children form and apply higher-order generalizations about the *kinds* of properties that are likely to be inductively rich. In fact, recent computational formulations of the "theory theory" of cognitive development have proposed that learners represent generalizations at multiple levels of abstraction, creating "overhypotheses" (Goodman, 1983/1955) that enable learners to learn quickly and generalize effectively to novel cases. Building on these ideas, the act of explaining could encourage children not only to favor properties that support broad generalizations in a given task, but also the *kinds* of properties that are typically reliable guides to particular inferences.

One open question – both in experimental and real-world contexts – relates to the role of pedagogical cues in fostering the benefits of explanation on inductive inference. Although the two-experimenter paradigm used in Experiment 1b ruled out certain pragmatic inferences that might have occurred as a direct result of the experimental procedure, children may have still interpreted the interactions pedagogically. Pedagogical learning does not necessarily require formal teaching, but rather a teacher's intent to communicate information to a learner in a context in which there exists some epistemic distance between those individuals (Shafto, Goodman, & Frank, 2012). Recent research suggests that children's interpretation of evidence may vary depending on whether learning occurs in pedagogical or non-pedagogical contexts (Bonawitz et al., 2011; Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Rhodes, Gelman, & Brickman, 2010; Shafto et al., 2012). In particular, previous research has shown that, like explanation, pedagogical cues can promote attention to inductively rich features (Csibra & Gergely, 2006; Csibra & Gergely, 2009). While the pedagogical cues in the current studies were well matched across *explain* and *control* conditions, it is certainly possible that *both* explanation *and* pedagogical cues may play a role in the effects reported here. The role of natural pedagogy in mediating or moderating effects of explanation on learning represents an important and novel avenue for future research.

6.1. Conclusions

Our data demonstrate that children as young as 3 years of age have the conceptual resources to reason on the basis of inductively rich properties, and that explanation facilitates their ability to avoid perceptually-bound judgments. In the current experiments, children had to decide whether to favor causal similarity or perceptual similarity in generalizing a hidden property or category membership from one object to another. Perceptual properties are often a reasonable basis for generalization, however, "insides" and category membership (labels) are more reliably associated with causal properties than with superficial, perceptual ones across many real-world cases. We propose that the process of explaining supports the construction and consultation of higher-order generalizations concerning such clusters of associated properties, in turn supporting inferences to new cases. By prompting children to favor inductively rich regularities, explanation encourages children to look beyond immediate observations to consider higher-order generalizations that support abstract knowledge.

Acknowledgments

This research was funded by the McDonnell Foundation and the National Science Foundation Grant BCS-331620 to A. Gopnik, and the McDonnell Foundation and National Science Foundation Grant DRL-1056712 to T. Lombrozo. The authors would like to thank the parents and children who participated in this research. We are also grateful to the University of California, Berkeley Early Childhood Centers, the Lawrence Hall of Science, Habitot, and the Center for Childhood Creativity at the Bay Area Discovery Museum for facilitating recruitment. Finally, we thank Rosie Aboody, Dustin Bainto, Ela Banerjee, Brynna Ledford, Donna Ni, and Gabriel Poon for assisting in data collection.

References

Amsterlaw, J., & Wellman, H. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7*, 139–172.

Bonawitz, E. B., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction affects spontaneous exploration and discovery. *Cognition, 120*, 322–330.

Booth, A. E., & Waxman, S. R. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology, 38*, 948–957.

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development, 82*, 2053–2066.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120*(3), 331–340.

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

Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.

Craik, K. J. W. (1943). *The nature of explanation* (pp. viii–123). Cambridge University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance XXI* (pp. 249–274). Oxford: Oxford University Press.

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month old infants. *Psychological Science, 21*, 1871–1877.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*(2), 164–168.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy, 71*, 5–19.

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

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a Dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*(5), 796–804.

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology, 66*, 327–353.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gelman, S., & Markman, E. (1987). Young children's induction from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532–1541.

Gelman, S., & Medin, D. M. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8*, 157–167.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*, 752–775.

Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory-formation system. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299–324). Cambridge, MA: MIT Press.

Gopnik, A., & Sobel, D. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 1205–1222.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18–32.

Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 411–422.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135–175.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversations. *Developmental Psychology, 37*(5), 668–683.

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology, 39*, 61–70.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*, 45–75.

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

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

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*, 221–236.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307–321.

Kirkham, N. Z., Cruess, L. M., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–467.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.

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

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development, 83*, 173–185.

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives, 8*, 101–106.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929–944.

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology, 126*, 198–212.

Legare, C. H., Wellman, H. M., & Gelman, S. A. (2009). Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology, 58*, 177–194.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464–470.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). Oxford, UK: Oxford Univ. Press.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition, 59*, 307–335.

Medin, D. M. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469–1481.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5*, 309–315.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science, 3*(4), 389–396.

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory and Cognition, 19*, 543–557.

Newman, G. E., Herrmann, P., Wynn, K., & Keil, F. C. (2008). Biases towards internal features in infants' reasoning about objects. *Cognition, 107*, 420–432.

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory and Cognition, 34*(1), 3–16.

Rhodes, M., Gelman, S. A., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science, 13*(3), 421–429.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge: Cambridge University Press.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science, 7*, 341–351.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition, 52*, 1–21.

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

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

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognitive Development, 8*(2), 159–182.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science, 25*, 161–169.

Walker, C. M., Lombrozo, T., Williams, J. J., & Gopnik, A. (submitted for publication). The role of explanation in children's causal learning.

Walker, C. M., Williams, J. J., Lombrozo, T., & Gopnik, A. (2012). Explaining influences children's reliance on evidence and prior knowledge in causal induction. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th ann conf of the cog sci soc* (pp. 1114–1119). Austin, TX: Cognitive Science Society.

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In L. Schulz, & A. Gopnik (Eds.), *Causal learning: Psychology, philosophy, & computation* (pp. 261–279).

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

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34*, 776–806.

Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology, 66*, 55–84.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*. http://dx.doi.org/10.1037/a0030996.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2011). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2011 ed.). <http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/>.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.