Developmental Psychology
2012, Vol. 48, No. 5, 1242–1253

0012-1649/12/$12.00 DOI: 10.1037/a0027540

Disagreement and Causal Learning: Others' Hypotheses Affect Children's Evaluations of Evidence

Andrew G. Young, Martha W. Alibali, and Charles W. Kalish
University of Wisconsin—Madison

When children evaluate evidence and make causal inferences, they are sensitive to the social context in which data are generated. This study investigated whether children learn more from evidence generated by an agent who agrees with them or from one who disagrees with them. Children in two age groups (5- and 6-year-olds and 9- and 10-year-olds) observed the functioning of a machine that lit up and played music in the presence of certain objects. After endorsing one of two plausible causal hypotheses, children observed a puppet either agree or disagree with their own hypotheses. The puppet then generated a further piece of evidence that confirmed, disconfirmed, or was neutral with respect to the children's hypotheses. When they were later asked to make causal inferences about objects they did not directly observe, children in both age groups responded differentially to identical evidence depending on whether the agent agreed or disagreed, and they often drew stronger inferences in response to disagreement. In addition, older children were particularly sensitive to disagreement when the evidence was ambiguous. Our results suggest that children consider the relationship between their own and others' hypotheses when evaluating evidence that others generate.

Keywords: causal inference, hypothesis testing, social learning, cognitive development

Data do not stand alone. Psychological studies of scientific and causal reasoning in adults and children have documented interdependence among evidence, theory, and experimentation. Many aspects of people's scientific and causal reasoning are affected by their prior knowledge and beliefs, such as their evaluations of evidence (Koslowski, 1996; Schulz, Bonawitz, & Griffiths, 2007), hypothesis formation (Schauble, 1996), and experimentation (Klahr, 2000). However, it is not only one's own knowledge and beliefs that may matter. Learning often happens in a social context, and therefore learners regularly encounter the beliefs, hypotheses, and experimental strategies of other agents. The present article explores the influence of others on children's causal learning.

This study explored how children use experimental evidence to draw conclusions. Experimental evidence is generated through interventions—direct manipulations of variables within a causal system (Woodward, 2003). Recent literature on the development of causal reasoning demonstrates that children can use interventions much like scientists do: to minimize confounding elements, identify spurious associations, and determine causal directionality (Gopnik et al., 2004; Schulz et al., 2007). This research has illuminated a surprising sophistication in children's ability to draw causal inferences that are based on interventions, prior beliefs, and complex statistical evidence (see Gopnik & Schulz, 2007). However, children do not explore the world alone. Psychologists have begun to consider the social processes involved in the generation and evaluation of evidence and theory (Feist, 2006). By most normative accounts of experimental logic, social information should not affect one's interpretation of experimental evidence (Butera & Mugny, 2001). Why, for example, should another's hypotheses affect one's interpretation of the evidence? Shouldn't the data speak for themselves? There is growing evidence that the answer is no.

Social processes may influence the cognitive mechanisms that underlie scientific and causal reasoning (Thagard, 1999). A developing body of evidence suggests that even very young children readily engage in social, intentional, and pragmatic reasoning to judge the outcomes of their own and others' causal interventions. The informativeness of another's interventions is potentially sub-

This article was published Online First March 5, 2012.

Andrew G. Young and Martha W. Alibali, Department of Psychology, University of Wisconsin—Madison; Charles W. Kalish, Department of Educational Psychology, University of Wisconsin—Madison.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Award R305C050055 to the University of Wisconsin—Madison; by Grant (DLS/DRM) 0745423 from the National Science Foundation to Charles W. Kalish; and by a Vilas Associate Award from the University of Wisconsin to Martha W. Alibali. The opinions expressed here are those of the authors and do not represent the views of the U.S. Department of Education or the National Science Foundation. Portions of this research were presented at the Biennial Conference of the Society for Research in Child Development (April 2009), the Biennial Meeting of the Cognitive Development Society (October 2009), and the 37th Carnegie Mellon Symposium on Cognition (October 2009).

Many thanks to the members of the Cognitive Development & Communication Lab and the Study of Children's Thinking Lab (both at the University of Wisconsin—Madison), to Timothy Rogers, and to Charles Snowdon for helpful discussions about this work and feedback on previous versions of the manuscript. We are especially grateful to the children, parents, and staff at Eagle's Wing Child Care and Education Programs at the University of Wisconsin—Madison, the University of Wisconsin Preschool Lab (Linden site), the Madison Central Montessori School, and the After School Program of the YMCA of Dane County.

Correspondence concerning this article should be addressed to Andrew G. Young, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, WI 53706. E-mail: agyoung2@wisc.edu

ject to that agent's own mental states, knowledge, skills, and abilities. For example, Kushnir, Wellman, and Gelman (2008) presented preschoolers with interventions generated by an expert puppet (i.e., "knows all about it") and a naïve puppet (i.e., "doesn't know anything about it"). Both puppets selected blocks and placed them simultaneously on a machine that subsequently activated (i.e., lit up and played music). Although both puppets' interventions were identically associated with the machine's activation, most children identified the expert puppet's block as the singular cause. Similar studies demonstrate that children distinguish between equivalent interventions performed intentionally and accidentally (Kushnir et al., 2008), by themselves and others (Kushnir, Wellman, & Gelman, 2009), and in the presence and absence of pedagogically informative rationales (Sobel & Sommerville, 2009). Thus, children's reasoning seems to be sensitive to a variety of aspects of the social context in which data-generating actions may be embedded.

A long tradition of research in social psychology demonstrates the pervasive effects of social factors on individual cognition and judgment. For example, people will endorse claims directly contradicted by objective evidence when a number of others make the same claims (Asch, 1956). More recently, reliance on others has been explored as a potentially rational strategy (Gelman, 2009). After all, other people are an important, and often critical, source of information about the world. Young children are quite discriminating with regard to whom they listen to or learn from. They productively track speakers' past reliability (Koenig & Harris, 2005), consider majority versus minority consensus (Corriveau, Fusaro, & Harris, 2009), weigh prior belief against speaker status/familiarity (Corriveau, Harris, et al., 2009), and consider speakers' expert causal knowledge (Sobel & Corriveau, 2010). Children clearly learn from others; what is less clear is how children integrate others' beliefs or claims with their own beliefs and objective experience. In the current study we held others' credibility constant and varied the relation between others' beliefs and children's own beliefs. Specifically, we examined whether agreement and disagreement between a child and another agent, in terms of their causal hypotheses, would affect the child's reasoning about interventions and their outcomes.

As social processes, agreement and disagreement are important contextual factors to consider for a number of reasons. Agreement and the formation of consensus seem to underlie many of the decisions scientific communities make in the interpretation of anomalous data (T. S. Kuhn, 1962; Thagard, 1999). Others' agreement is a relevant index of the prior probability of a hypothesis. On the other hand, disagreement is an important potential source of evidence counter to one's favored theory. First, disagreement suggests the possibility that one is mistaken and highlights potential alternative theories (Christensen, 2009). Second, disagreement is often accompanied by the presentation of anomalous data. Although anomalous data presented by dissenters can often be forcefully ignored, discounted, and distorted for a surprising length of time (e.g., Oreskes, 1999; Thagard, 1999), disagreement nonetheless seems to be nearly ubiquitous and crucial to the generation of scientific knowledge (Silverman, 1992).

Disagreement and counterevidence may play a similar role in children's theory change and cognitive development (Gopnik & Wellman, 1992, 1994). Many researchers posit sociocognitive conflict (i.e., disagreement) as a primary mechanism of conceptual change in science learning (e.g., Ames & Murray, 1982; Inhelder & Piaget, 1958; Schwarz, Neuman, & Biezuner, 2000; Williams & Tolmie, 2000). For example, Howe (2009) found that 8- to 12-year-olds showed greater conceptual gains concerning physical systems (e.g., motion down inclined planes and object buoyancy) in collaborative groups in which members focused on their conflicting prior beliefs than in groups that emphasized resolution of disagreements. Furthermore, strength of disagreement was positively associated with conceptual gains. Thus, agreement and disagreement seem to have real consequences for learners.

Despite the literature on sociocognitive conflict and conceptual change, very little is known about children's responses to agreement or disagreement when evaluating evidence. Thus, at the broadest level, our research goal was to establish whether an intervening agent's agreement or disagreement with a child's causal hypotheses would affect subsequent causal learning from evidence. It may be an effective (if not optimal and normative) strategy for learners to pay equal attention to the outcomes of all interventions regardless of an agent's beliefs and agreement. However, it is also possible that the agent's beliefs might affect the informativeness of the evidence.

How or why might agreement and disagreement affect children's reasoning? One possibility is that children might treat an agent's hypothesis as an index of collaboration. Research suggests that children learn better from others when they are engaged in joint activity (e.g., Sommerville & Hammond, 2007; Tomasello, Kruger, & Ratner, 1993). Agreement may support the perception of collaboration, while disagreement may hinder it, which suggests that children would learn more from agreeing agents than disagreeing agents. Alternatively, disagreement may provoke the child to consider alternative hypotheses. An agreeing agent may lead to overconfidence, whereas a disagreeing agent may spur the child to pay closer attention to evidence. This possibility suggests that children may learn more from disagreeing agents, especially when initial hypotheses are false. Finally, recent thinking about the normative implications of agreement and disagreement highlights that agreement should matter most in circumstances of ambiguity (Christensen, 2009). If this is the case, children might consider others' hypotheses only when circumstances are ambiguous or in need of a "tie breaker."

In the present study, we assessed children's causal learning from another's interventions in two age groups: 5- and 6-year-olds and 9- and 10-year-olds. From early to middle childhood, there are numerous changes in inductive reasoning, evidence evaluation skills, and epistemic understanding. In general, children's learning from ambiguous and disconfirming evidence increases in sophistication over the elementary school years (Zimmerman, 2007). Children also become better able to differentiate and coordinate theory and evidence over this age span (D. Kuhn, 1989; Zimmerman, 2007). Furthermore, children's awareness that two people might form different beliefs even though they use the same evidence seems to emerge around age 7 (Chandler & Lalonde, 1996), as does their appreciation of inference (as opposed to direct observation) as a warrant for hypotheses (Ruffman, Perner, Olson, & Doherty, 1993). In light of these findings, we expected that children in these younger and older age groups would respond differently to agreement and disagreement. In particular, we expected that younger children would either ignore others' hypotheses or treat them as data akin to direct observations. Five and 6-year-olds often believe that if two people disagree, only one can be correct (Chandler & Lalonde, 1996; Lalonde & Chandler, 2002), so they may view support for one hypothesis as undermining the other. Nine and 10-year-olds, in contrast, appreciate different qualities of evidence and may recognize that support for one hypothesis does not necessarily undermine all others. Thus, we expected that older children would be most sensitive to others' hypotheses when the data were ambiguous.

We chose to use a naïve puppet as the collaborating agent in this study. Several factors informed this decision. First, there is reason to believe that young children will treat a naïve puppet as an epistemic peer more so than they will an adult confederate. Children tend to treat adults' beliefs as reliable and generally accurate by default (e.g., Corriveau, Harris, et al, 2009; Jaswal, 2010). Second, children in causal learning studies readily defer to agents with domain-specific causal knowledge, even when those agents are puppets (e.g., Kushnir et al., 2008). By using a naïve puppet, we limited the inherent "weight" of the collaborating agent's beliefs, providing a neutral context within which to examine our primary research question of how an agent's agreement or disagreement influences children's reasoning.

Using a modified "detector" task (e.g., Gopnik et al., 2004; see Figure 1), we had children view demonstrations of two objects activating and failing to activate a detector, such that there were two viable hypotheses as to which object feature was causally responsible. Children then endorsed one of the two hypotheses and sorted objects according to their expected causal powers. Next, we introduced a puppet named Leo (i.e., the intervening agent). We varied whether the puppet endorsed the participant's chosen hypothesis (*agree*) or the counterhypothesis (*disagree*) prior to asking the puppet to generate a third piece of evidence. We also included a control condition in which the puppet did not explicitly state a hypothesis. We also varied whether the evidence the puppet generated *confirmed*, *disconfirmed*, or was *neutral* with respect to the child's hypothesis (see Table 1). Following the puppet's intervention, children once again sorted the objects according to their causal powers to activate the detector.

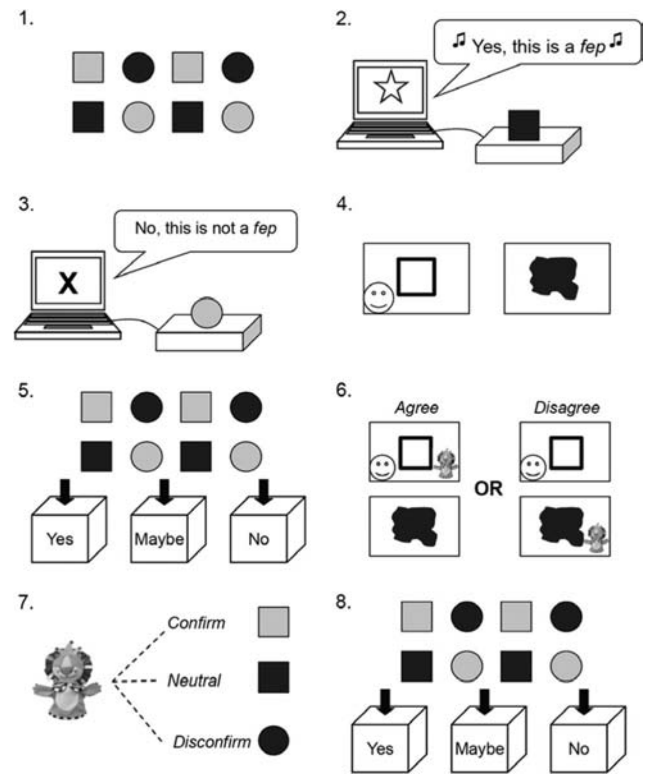

Figure 1. Summary of procedure (shape colors red and blue are represented by dark grey and light grey, respectively): (1) An object set is presented. (2) Child selects an object ($A+B+$) that activates detector. (3) Experimenter selects opposite-featured object ($A-B-$) that fails to activate detector. (4) Child endorses a hypothesis. (5) Child sorts objects according to their ability to activate detector. (6) Leo selects a hypothesis (agent hypothesis: agree or disagree). (7) Leo selects an additional object that activates the detector (agent evidence: confirming, neutral, or disconfirming). (8) Child sorts objects according to their ability to activate detector.

Method

Participants

Forty-five 5- and 6-year-olds ($M = 6$ years, 0 months; $SD = 6.5$ months; 25 girls) and thirty-six 9- and 10-year-olds ($M = 9$ years, 9 months; $SD = 7.8$ months; 24 girls) were recruited from early-childhood centers and after-school programs serving a largely middle-class population in a mid-sized Midwestern city. Children in each age group were randomly assigned to an experimental or a control condition by a ratio of 2:1, with the constraint that ages had to be roughly equal across the conditions. Thus, 30 younger and 24 older children participated in the experimental condition, and 15 younger children and 12 older children participated in the control condition.

Materials

A laptop computer and sham detector attachment were used to generate evidence. The detector attachment consisted of a 20 × 16 × 8 cm plastic and metal enclosure that housed flashing LEDs. Objects were placed on the (nonfunctional) detector attachment throughout the study. On the computer, a graphic user interface operated on prescripted sequences to generate evidence. Within the interface, the experimenter could turn a dial to set the detector to "look" for different kinds of objects (e.g., *feps, daxes, blickets*) and initiate detection sequences. When "activated," the interface played music, became highly animated, and vocally confirmed the object as the source of activation (e.g., "Yes, this is a *fep*!"). When an object "failed to activate" the detector, the interface played an abrupt error noise, produced a large red X, and vocally confirmed the object's failure (e.g., "No, this is not a *fep*!"). This apparatus provides a close analog to several features of the detectors used by Gopnik, Sobel, and colleagues (e.g., Gopnik et al., 2004).

Children completed one practice trial and six experimental trials, each involving a different novel object set. Each set of objects had two distinct binary feature dimensions (e.g., the practice set colors were red and blue, and the shapes were cubes and spheres), so there were a total of four unique objects within a set. Full object sets contained two instances of each object kind. Experimental object set features were chosen to be salient and plausible causes of activation (for details and examples, see the

Table 1
Summary of Study Design and Conditions

Participants	Agent hypothesis	Agent evidence		
		Confirming	Neutral	Disconfirming
Experimental group (30 younger, 24 older)	Agree	1 trial	1 trial	1 trial
	Disagree	1 trial	1 trial	1 trial
Control group (15 younger, 12 older)	None	2 trials	2 trials	2 trials

Appendix). For each object set, four hypothesis cards illustrating each possible feature value were used to represent possible hypotheses to children.

Other materials included the collaborating agent Leo, a children's lion puppet, and three clear plastic sorting boxes marked "Yes," "Maybe," and "No." Miscellaneous materials included a marker and a crayon for demonstration purposes and plastic tokens for the children and Leo to use to indicate their hypotheses.

Procedure

Children participated individually with an experimenter in a quiet place at their school or in our laboratory. Each session was videotaped. The child sat beside the experimenter with the detector and sorting boxes in front of him or her. The experimenter introduced the task by saying, "Today we are going to play a game, and the point of this game is to learn about some toys. But, these toys are very strange! So I brought this machine to help us!" The experimenter then familiarized the child with the operations of the detector. To demonstrate the machine's abilities, the experimenter set the detector to look for a familiar object, a marker, by turning a dial on the machine. The child then tested a marker and a crayon. The experimenter elaborated upon the detector's functioning while the marker and the crayon, respectively, activated and failed to activate the detector.

Following familiarization, the experimenter introduced Leo along with a cover story. The participant was told that Leo desired a job at a toy factory but needed help because he didn't know anything about the toys (e.g., what they look like or their names). The experimenter then initiated a practice trial with an object set consisting of red and blue cubes and spheres (see Figure 1 for a summary of the procedure). The experimenter placed the objects on the table and said, "Some of these toys are *feps*, and some of these toys are not *feps*. Do you know what a *fep* is? I don't either! Our job and Leo's job is to figure out which kinds of toys are *feps*. Let's use the machine to find out." The experimenter then set the detector to look for *feps* and asked the child to pick an object to test. The child then chose an object and placed it on the detector. The child's choice always activated the detector. This first object is referred to as the $A+B+$ object, where A and B refer to the object's two feature dimensions. The A feature was whichever dimension the child later endorsed as the cause of the activation. The B feature was the alternative feature, deemed noncausal by the child (see below). The experimenter then set the $A+B+$ object aside and selected another object to be tested. The object selected was always that with the opposite feature levels of $A+B+$. That is, if $A+B+$ was a red cube, then the experimenter selected a blue sphere. The experimenter's choice never activated the detector. This second object is referred to as the $A-B-$ object.

Following the initial testing of the two objects, the experimenter introduced the child to two plausible causal hypotheses. The experimenter first restated that the $A+B+$ object was a *fep* and that the $A-B-$ object was not a *fep*. Then the experimenter directed attention to the two features of the $A+B+$ object (e.g., red and cube) in random order by saying, "This toy is a *fep*. It is a red toy and it is a square toy. That means *feps* could be red toys or *feps* could be square toys." At this time, the experimenter presented two hypothesis cards corresponding to the features of the $A+B+$ object and asked, "What kind of toys do you think are *feps?* Red toys or square toys? If you think red toys are *feps*, you should put your token on the red card. If you think square toys are *feps*, you should put your token on the square card. Go ahead and make your best guess." The participant's selection is hereinafter referred to as the *participant hypothesis*. In addition, the feature selected by the participant is always designated as the A feature. Thus, the participant's hypothesis is always that the A feature caused the detector to activate. The nonselected feature is always designated as the B feature.

Once the child endorsed a hypothesis, he or she was asked to sort the objects. The experimenter directed the child by saying, "Now it's time to put these toys where they belong. This box says 'Yes'; it's for *feps*. You should put the toys that are *feps* in here. This box says 'No'; it's for toys that aren't *feps*. You should put the toys that are not *feps* in here. This middle box says 'Maybe'; it's for toys you're not sure about. You should put toys that might be *feps* or toys you don't know about in here." The practice trial concluded once a child had sorted all the objects into the boxes.

Following the practice trial, children completed six test trials, each with a different object set and target label (for illustrative purposes, in describing the task we continue to use the practice set label *fep* and the color/shape features). The initial phase of each test trial proceeded exactly as the practice trial above but with less experimenter direction and with less reference to the cover story. The child was shown a novel object set, the detector was set to detect a novel target toy (e.g, *daxes*), the child chose an activating object (deemed the $A+B+$ object), the experimenter chose a nonactivating object (deemed the $A-B-$ object), the child endorsed a hypothesis, and finally, the child sorted the objects for the first time (which we refer to as the *presort*). A second phase of each test trial gave Leo the opportunity to select a hypothesis and test an object.

The experimenter began the second phase by saying, "Let's give Leo a turn!" The objects were removed from the sorting boxes and Leo was removed from a container outside the view of the task area. For children in the experimental condition, Leo was first asked to endorse a hypothesis, "Leo, it looks like [child's name] thinks square toys are *feps*. What kind of toys do you think are *feps*?" Operated by the experimenter, Leo then selected one of the two possible hypotheses by placing his token on the appropriate card. Leo's hypothesis choice, which we refer to as the *agent hypothesis*, was experimentally manipulated. Leo could *agree* with the participant and select the participant hypothesis (i.e., *A* feature is causal) or *disagree* and select the counterhypothesis (i.e., *B* feature is causal). The experimenter then made verbal note of the agreement ("Looks like Leo thinks square toys are *feps* too!") or disagreement ("Looks like Leo thinks red toys are *feps*, **not** [with emphasis] square toys.").

Following Leo's hypothesis selection, the experimenter invited Leo to test an object by saying, "Leo, go ahead and try a toy!" Across trials Leo provided three types of tests, or *agent evidence*: (a) *confirming*—Leo activated the detector with the remaining object that was consistent with the participant hypothesis (i.e., the $A+B-$ object); (b) *disconfirming*—Leo activated the detector with the remaining object that was consistent with the counterhypothesis (i.e., the $A-B+$ object); and (c) *neutral*—Leo activated the detector with the originally tested object (i.e., the $A+B+$ object). Note that Leo always activated the detector and never tested the nonactivating $A-B-$ object. The experimental trial concluded with children sorting the objects into the Yes, Maybe, and No boxes a final time (which we refer to as the *postsort*).

Children in the control condition completed test trials identical to those in the experimental condition with the exception that Leo was not asked to select a hypothesis. Following a child's first sort, Leo was introduced and immediately asked to try a toy. Thus, the agent expressed no hypothesis in the control condition.

Children in both conditions completed six test trials (see Table 1). Children in the experimental group completed one trial under each combination of agent hypothesis (*disagree, agree*) and agent evidence (*confirming, disconfirming, neutral*). Children in the control group completed two trials at each level of agent evidence in the absence of any agent hypothesis. The three levels of agent evidence (*confirming*, *disconfirming*, *neutral*) appeared in counterbalanced order across participants. Trials were blocked by agent hypothesis (*agree*, *disagree*) in the experimental condition. Object sets were randomly assigned to trials.

Children's object sorts were coded to provide the dependent measure. Individual objects received scores of 2, 1, or 0 corresponding to their placement in the Yes, Maybe, or No sorting boxes. These scores were aggregated across both instances of each object, resulting in pre- and postsort scores for each type of object ($A+B+$, $A+B-$, $A-B+$, and $A-B-$) which ranged from 0 to 4.

Results

Our analyses investigated the effects of the agent's hypothesis on children's responses to evidence. In the first section, we evaluate children's sorting of the objects before the agent (Leo) advanced a hypothesis and generated evidence (i.e., their presorts). In the second section, we examine changes in children's sorting after Leo advanced his hypothesis and generated evidence. In the third section, we briefly examine how children's postsorts varied in response to different types of agent evidence. These three sections consider only children in the experimental condition. Thus, in these sections we directly evaluate whether agent hypothesis affects children's responses to evidence. In the fourth section, we briefly revisit our major findings involving agent hypothesis by comparing children in the experimental group to those in the control group, in which the agent advanced no hypotheses.

We found no effects of child gender, order of agent evidence, or testing location (lab vs. school) in any of the analyses, so we collapsed these factors. Furthermore, we found no effect of object set on any outcome measures. Order of agent hypothesis (i.e., whether Leo agreed or disagreed on the first three trials) significantly affected children's initial sorts of the $A+B-$ and $A-B+$ objects; however, the differences associated with order were not in direction but only in magnitude. That is, the overall trends in the presort data held at each order of agent hypothesis. Furthermore, agent hypothesis order was not significant as a main effect or in interaction with other factors in any other analyses. Thus, for simplicity, we do not report specific order effects.

Did Children Initially Sort in Accordance With Their Endorsed Hypotheses?

We first examined whether children understood the task and responded according to their endorsed hypotheses. Children in both age groups did so. Figure 2 presents children's sorting scores after the initial evidence phase across the six test trials. Consistent with the initial evidence, participants sorted $A+B+$ objects as most likely to activate the detector and $A-B-$ objects as least likely to activate the detector. Children also sorted the remaining, untested objects according to their endorsed hypotheses. By definition, the $A+B-$ object was consistent with the child's hypothesis and the $A-B+$ object was inconsistent. Children sorted $A+B-$ objects as being more likely to activate the detector than were $A-B+$ objects. Pair-wise comparisons revealed significant differences among presort scores for all pairs of object types ($ps <$ .001). Children's presorts were equivalent across agent hypothesis and agent evidence conditions, which was expected because the presorts occurred prior to Leo's actions.

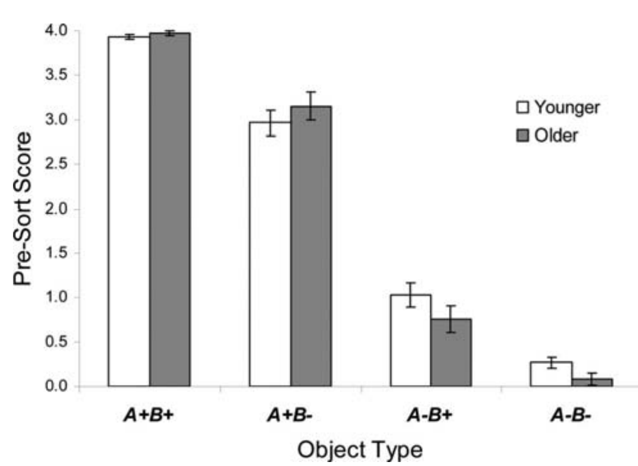

Figure 2. Mean presort score by object type and age group. Error bars reflect standard errors.

Did Children's Responses Change Depending on the Agent's Hypothesis and Evidence?

We next examined whether children's sorting changed in response to whether Leo agreed or disagreed with their hypothesis and generated evidence that was confirming, disconfirming, or neutral. To do so, we utilized change scores that reflected children's increased or decreased expectations of objects' causal powers to activate the detector. These change scores were derived by subtracting each object's presort score from its postsort score. Sorting an object as No at presort and Yes at postsort would result in a change score of 2; conversely, sorting an object as Yes at presort and No at postsort would result in a change score of −2. Because there were two instances of each object, change scores for each object type ranged from −4 to 4.[1]

We constructed an overall *hypothesis-change score* by subtracting the change scores for the *A*−*B*+ object from those for the *A*+*B*− object. These scores reflect the extent to which children's sorting of the previously untested *A*+*B*− and *A*−*B*+ objects became more or less consistent with their endorsed hypotheses. Scores ranged from −8 to 8.

Hypothesis change. Did Leo's agreement or disagreement affect children's responses to evidence? The answer is yes. Figure 3 presents children's hypothesis-change scores as a function of Leo's agreement or disagreement and the nature of the evidence Leo generated. We analyzed these data using analysis of variance (ANOVA), with agent hypothesis (agree, disagree) and agent evidence (confirming, disconfirming, neutral) as within-subject variables, age group as a between-subjects variable, and hypothesis-change score as the dependent variable.

Across both age groups, children's sorting varied as a function of Leo's evidence, $F(2, 104) = 113.89$, $p < .001$, $\eta_p^2 = .687$. Furthermore, children's sorts became less consistent with their original hypotheses when Leo disagreed than when he agreed, $F(1, 52) = 10.75$, $p = .002$, $\eta_p^2 = .171$. In addition to these main effects of agent evidence and agent hypothesis, there was also a three-way interaction of agent evidence, agent hypothesis, and age group, $F(2, 104) = 20.69$, $p = .004$, $\eta_p^2 = .101$. As can be seen in Figure 3, younger children showed the strongest response to Leo's disagreement when evidence was disconfirming, whereas older children showed an exaggerated response to Leo's disagreement when evidence was neutral. The simple effect of agent agreement approached significance for disconfirming evidence in younger children, $F(1, 29) = 4.07$, $p = .053$, $\eta_p^2 = .123$, and was significant for neutral evidence in older children, $F(1, 23) = 14.57$, $p = .001$, $\eta_p^2 = .388$.

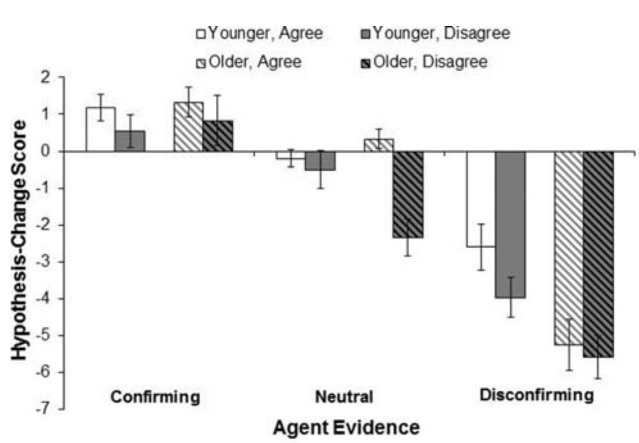

Figure 3. Mean hypothesis-change score by agent evidence, agent hypothesis, and age group. Error bars reflect standard errors.

These findings provide preliminary evidence that Leo's hypotheses influenced children's overall sorts. However, hypothesis-change scores are a fairly coarse measure of children's specific responses. As a composite, hypothesis-change scores may obscure systematic responses for individual objects. Furthermore, systematic responses for individual objects could potentially negate each other in the overall hypothesis-change scores. Thus, we next examined change scores for individual objects.

Note that at the children's final sort, individual objects had different evidential status. Some objects were tested before children endorsed a hypothesis, some were tested by Leo, and some were not tested at all. For the tested objects, children's sorting reflected their memory of whether the object did or did not activate the detector. For those objects not tested, children's sorting reflected an inference or prediction. It is important to note that the objects Leo tested varied by agent evidence condition. Specifically, Leo tested the *A*+*B*− object in the confirming condition and the *A*−*B*+ object in the disconfirming condition. Thus, simply comparing sorts of the same object across conditions would conflate very different kinds of evidence. For this reason, we first present results for objects not tested during the trials and then consider objects for which children had direct evidence.

Not-tested objects. Before Leo's actions, two objects had never been observed to interact with the detector: the *A*−*B*+ and *A*+*B*− objects. In the confirming condition, Leo found that the *A*+*B*− object activated the detector, implying that the *A*−*B*+ object did not. In the disconfirming condition, Leo found that the *A*−*B*+ object activated the detector, implying that the *A*+*B*− object did not. In each case, the evidence suggested that the not-tested object would not activate the detector. Thus, we expected change scores for the not-tested objects to be negative. Our main question, however, was whether this direct effect of evidence would depend on agent hypothesis. (Note that the neutral evidence condition left two objects untested. This condition is considered separately.)

To address these questions, we conducted an ANOVA with agent hypothesis (agree, disagree) and agent evidence (confirming, disconfirming) as within-subject variables and age group as a between-subjects variable. The dependent measure was the change score for the not-tested object (i.e., *A*−*B*+ change in the confirming condition and *A*+*B*− change in the disconfirming condition). As can be seen in Figure 4, there was a significant interaction between agent hypothesis and agent evidence, $F(1, 52) = 6.82$, $p = .012$, $\eta_p^2 = .116$. When Leo produced confirming evidence, there was no effect of agent hypothesis. However, when Leo

[1] Note that we do not refer to change scores for the *A*+*B*+ and *A*−*B*− objects in the following analyses. Given the direct evidence children generated for the *A*+*B*+ and *A*−*B*− objects prior to Leo's interventions, no significant change in sorting for these objects was observed in response to manipulations of agent hypothesis or agent evidence. Thus, we do not consider them further.

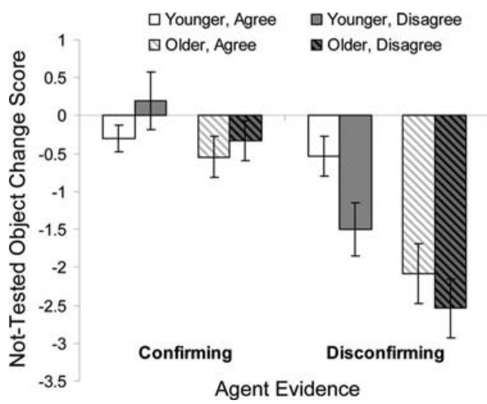

Figure 4. Mean not-tested object change score by agent evidence, agent hypothesis, and age group. Error bars reflect standard errors.

produced disconfirming evidence, there was a simple effect of agent hypothesis (agree, disagree). When Leo disagreed, children of both ages were more likely to appreciate that the $A{+}B{-}$ object would not activate the detector, $F(1, 52) = 6.55$, $p = .013$, $\eta_p^2 = .112$. That is, children were more accurate in accommodating evidence produced by a disagreeing agent.

This effect was also evident when the data were analyzed using a nonparametric test. Children's distributions of responses to disconfirming evidence varied by agent hypothesis, Stuart-Maxwell test, $\chi^2(2, N = 54) = 7.24$, $p = .027$. Children were more likely to have negative change scores (61% vs. 48%) and less likely to have positive change scores (4% vs. 17%) when Leo disagreed than when he agreed.

In addition, a main effect of age group, $F(1, 52) = 9.88$, $p = .003$, $\eta_p^2 = .160$, revealed that older children changed their sorts more than did younger children, which suggests a greater understanding of the consequences of Leo's intervention.

In the neutral agent evidence condition, Leo's testing of the $A{+}B{+}$ object left both the $A{+}B{-}$ and $A{-}B{+}$ objects untested, providing no change in support for either hypothesis. Under these circumstances, did Leo's agreement or disagreement affect children's beliefs? To find out, we conducted two separate ANOVAs with agent hypothesis (agree, disagree) as a within-subject variable and age group as a between-subjects variable, with change scores for the $A{+}B{-}$ and $A{-}B{+}$ objects in the neutral condition as the dependent variables. The data are presented in Figure 5.

We first consider the $A{+}B{-}$ object (left set of bars in Figure 5). Older children viewed this object as less likely to activate the detector when Leo disagreed with them; that is, they changed their sorting to align more with Leo. Younger children did not reliably change their sorting of the $A{+}B{-}$ object when faced with neutral evidence regardless of Leo's hypothesis. This pattern yielded an agent hypothesis by age group interaction, $F(1, 52) = 12.43$, $p = .001$, $\eta_p^2 = .193$. Follow-up tests at each age group revealed a simple effect of agent hypothesis for older children, $F(1, 23) = 15.76$, $p < .001$, $\eta_p^2 = .407$, but not for younger children.

This effect was also evident when the data were analyzed using a nonparametric test. As in the ANOVA analysis, older children's overall distributions of responses varied by agent hypothesis, Stuart-Maxwell test, $\chi^2(2, N = 24) = 11.38$, $p = .003$. Older children were more likely to have negative change scores (48% vs. 0%) and less likely to not change (i.e., 48% vs. 83%) when Leo disagreed than when Leo agreed.

We found similar results for the $A{-}B{+}$ object (right set of bars in Figure 5). Younger children tended to maintain their belief that the $A{-}B{+}$ object would not activate the detector regardless of whether Leo agreed or disagreed; thus, the simple effect of agent hypothesis was not significant for the younger children. However, older children were more likely to indicate that the $A{-}B{+}$ object would activate the detector when Leo disagreed than when he agreed, yielding a simple effect of agent hypothesis for older children, $F(1, 23) = 10.15$, $p = .004$, $\eta_p^2 = .306$. This pattern was also observed when older children's responses were analyzed using a nonparametric test. Older children were more likely to have positive change scores (42% vs. 4%) and less likely not to change (46% to 88%) when Leo disagreed than when Leo agreed, Stuart-Maxwell test, $\chi^2(2, N = 24) = 10.00$, $p = .003$.

Thus, when presented with neutral evidence, older children became less committed to their initial hypotheses when Leo disagreed with them. Their most frequent response to Leo's disagreement was to sort the $A{+}B{-}$ and $A{-}B{+}$ objects into the Maybe box. However, when Leo agreed with them, older children did not become more committed to their hypotheses. It is important to note that Leo's disagreement did not cause younger children to become less committed to their hypotheses.

Tested objects. How did Leo's actions affect children's responses to objects that had directly interacted with the detector? In the confirming and disconfirming agent evidence conditions, children observed Leo test the $A{+}B{-}$ and $A{-}B{+}$ objects (respectively). Did agent hypothesis affect children's evaluations of these tested objects? To find out, we conducted an ANOVA with agent hypothesis (agree, disagree) and agent evidence (confirming, disconfirming) as within-subject variables and age group as a between-subjects variable, with change scores for the tested objects as the dependent measure (i.e., scores for the $A{+}B{-}$ object in the confirming condition and the $A{-}B{+}$ object in the disconfirming condition). No effect of agent hypothesis was found, suggesting that children readily learned about the causal status of the objects Leo tested regardless of whether he agreed or disagreed.

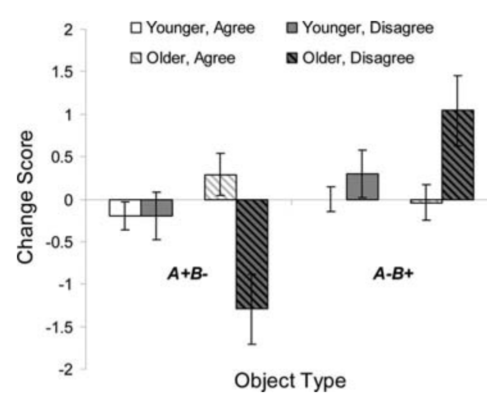

Figure 5. Mean $A{+}B{-}$ and $A{-}B{+}$ object change scores by agent hypothesis and age group in response to neutral agent evidence. Error bars reflect standard errors.

Agent Evidence and Children's Final Sorts

A closer look at children's postsorts provides more direct information on the effects of agent evidence. As can be seen in Figure 4, children changed their sorting of untested objects more in response to disconfirming evidence than confirming evidence ($p < .001$). Given children's differential presorts for the $A+B-$ and $A-B+$ objects (see Figure 2), this result is not unexpected, as children had more room and reason to change their beliefs about the $A+B-$ object in response to the disconfirming evidence. However, an effect of agent evidence remained even in children's final sorting of the not-tested objects. Younger children judged the not-tested object to be less likely to activate the detector when Leo provided confirming evidence ($M = 0.98$, $SE = 0.17$) than when he provided disconfirming evidence ($M = 1.85$, $SE = 0.23$), $F(1, 29) = 22.53$, $p = .003$, $\eta_p^2 = .273$. Similarly, older children judged the not-tested object to be less likely to activate the detector in response to confirming evidence ($M = 0.40$, $SE = 0.15$) than in response to disconfirming evidence ($M = 0.75$, $SE = 0.22$), $F(1, 23) = 7.50$, $p = .012$, $\eta_p^2 = .246$. Thus, regardless of age or agent hypothesis, children were better able to appreciate the noncausal status of the not-tested object in response to confirming evidence.

Similar patterns were observed for children's postsorts of the objects Leo tested. Younger children's final sorting more accurately reflected the causal powers of the tested objects in response to confirming evidence ($M = 3.78$, $SE = 0.10$) than in response to disconfirming evidence ($M = 3.25$, $SE = 0.18$), $F(1, 29) = 6.52$, $p = .016$, $\eta_p^2 = .184$. Older children's postsorts of tested objects did not depend on agent evidence ($p > .1$), reflecting a better ability to integrate their inaccurate previous belief with observed disconfirming data.

Together, these results echo previous research documenting children's general difficulty with disconfirming evidence. Furthermore, children's differential responses to disconfirming and confirming evidence suggest that children in both age groups committed to their endorsed hypotheses beyond just a surface level.

Control Comparisons

Thus far we have observed several key differences in children's evaluations of evidence depending on whether Leo agreed or disagreed with their hypotheses. Knowing how children would respond to the task if Leo did not advance any hypothesis would allow us to gauge whether Leo's disagreement facilitated evidence evaluation or whether his agreement hindered it. To address this question, we analyzed data from the first three trials of the 54 experimental and 27 control participants, using agent hypothesis (agree, disagree, none) as a between-subjects factor. There were no presort differences across agent evidence conditions, agent hypothesis conditions, or age group in this full sample.

We first examined children's change scores for the not-tested objects in response to confirming and disconfirming evidence when Leo advanced no hypothesis. As in the within-subjects analysis of the experimental group, there was an agent hypothesis by agent evidence interaction, $F(2, 75) = 3.60$, $p = .032$, $\eta_p^2 = .088$. Paralleling the earlier analysis, for confirming evidence, children changed their sorts relatively little, and Leo's hypotheses (agree, disagree, none) did not differentially affect their change scores. In response to disconfirming evidence, children more dramatically changed their sorts. However, despite the significant overall interaction, the simple effect of agent hypothesis (agree, disagree, none) for disconfirming evidence was not found (as it was in the higher powered within-subject analysis of the experimental group). Across both age groups, change scores in the control (i.e., none) condition ($M = -2.00$, $SE = 0.37$) generally fell between those observed when Leo agreed ($M = -1.26$, $SE = 0.41$) and disagreed ($M = -2.22$, $SE = 0.40$). It thus appears that there may be a mixture of Leo's disagreement facilitating evidence evaluation and Leo's agreement hindering it.

We also examined children's responses to neutral evidence when Leo did not advance a hypothesis. To do so, we conducted separate ANOVAs with agent hypothesis and age group as between-subjects variables and with $A+B-$ and $A-B+$ change scores as the dependent variables. We first consider the $A+B-$ object. Consistent with the within-subject findings, Leo's hypothesis affected only older children's sorts, as revealed in an agent hypothesis by age group interaction, $F(2, 75) = 10.43$, $p < .001$, $\eta_p^2 = .218$, and a simple effect of agent hypothesis for older children, $F(2, 33) = 7.98$, $p = .001$, $\eta_p^2 = .326$, but not for younger children. Post hoc comparisons showed that older children's $A+B-$ change scores were more negative when Leo disagreed ($M = -1.58$, $SE = 0.54$) than when Leo either agreed ($M = 0.50$, $SE = 0.26$) or had no hypothesis ($M = -0.33$, $SE = 0.23$), $ps < .01$. Similarly, for the $A-B+$ object, agent hypothesis affected children's sorts, $F(2, 75) = 4.09$, $p = .021$, $\eta_p^2 = .098$. However, given the lower power of the between-subjects comparison, this result did not hold for older children alone as a simple effect. Over both age groups, change scores for the $A-B+$ object were more positive when Leo disagreed ($M = 0.56$, $SE = 0.26$) than when he agreed ($M = -0.07$, $SE = 0.17$) or when he advanced no hypothesis ($M = -0.37$, $SE = 0.26$), $ps < .05$.

In sum, for both the $A+B-$ and $A-B+$ objects, children largely maintained their causal beliefs in the presence of neutral evidence both when Leo agreed and when Leo did not offer a hypothesis. Therefore, it appears that disagreement is responsible for children's more accurate recognition of causal indeterminacy in the face of uninformative evidence.

Discussion

The primary aim of this study was to determine whether children consider others' hypotheses in relation to their own when evaluating experimental evidence. Our results suggest that children do attend to and consider others' hypotheses. Broadly speaking, children's beliefs tended to more accurately reflect the state of affairs in response to evidence produced by agents who disagreed with them.

It is important to note that disagreement affected children's reasoning when they were faced with disconfirming or neutral evidence but not confirming evidence. Furthermore, the agent's hypothesis did not affect children's beliefs about objects for which they had direct evidence (i.e., $A+B+$, $A-B-$, and the object Leo tested). However, children's inferences about objects for which they had no direct evidence were sensitive to the agent's hypotheses. These limited conditions suggest a rather sensible approach to using information about others' hypotheses: Children responded to an agent's disagreement only in situations in which there was some degree of ambiguity regarding the correct causal judgment.

Across the inductive judgments our task presented to children, a disagreeing agent elicited stronger responses than an agreeing agent. In the disconfirming and confirming evidence conditions, the agent's tested object successfully activated the detector, implying that the remaining untested object was noncausal. If children's beliefs regarding the not-tested object were initially inaccurate (i.e., agent evidence was disconfirming), then a disagreeing agent helped them correctly revise their hypotheses more so than did an agreeing agent. When agent evidence confirmed the children's hypotheses, whether the agent agreed or disagreed did not cause them to change their hypotheses. However, this null result may be due to a ceiling effect, as children sorted according to their hypotheses on their initial sort and had little room for improvement.

With neutral agent evidence, neither hypothesis was confirmed or disconfirmed. Two objects remained untested, and the identity of the correct causal feature remained unknown. Under these conditions, younger children maintained similar beliefs from pre- to postsort regardless of agent hypothesis (i.e., $A+B-$ is causal and $A-B+$ is noncausal). However, agent disagreement, but not agreement, led older children to adjust their causal judgments. In particular, older children became less confident about the untested objects, more frequently placing them into the "Maybe" box. Furthermore, comparisons with the control condition, in which the agent advanced no hypothesis, revealed that it was disagreement that led to change, not agreement that led to maintainence. The presence of this effect in older, but not younger, children is the primary developmental difference we observed in response to the agent hypothesis manipulation.

One possible explanation for this developmental difference is that younger children had more difficulty recognizing the complete causal ambiguity posed by the neutral evidence condition. Chandler and Lalonde (1996) argued that 5-year-olds (and 6-year-olds to an extent) have difficulty accepting that the same evidence may be consistent with opposing beliefs (e.g., children tend to think, "If I am right, you must be wrong"). Our neutral evidence condition provided just such a situation. The 5- and 6-year-olds might have felt that since the result of the test was consistent with their hypothesis it must, ipso facto, be inconsistent with the agent's counterhypothesis. Older children likely appreciated the ambiguity of the neutral evidence, at least when the agent's disagreement called their attention to the fact that there was a viable alternative. This suggests that, with development, young children gradually come to appreciate the intermediate steps between firm knowledge and mere guesses. We speculate that increasing experience with evidence counter to one's own and others' hypotheses facilitates both recognizing ambiguity and understanding gradations of support for hypotheses.

Our findings clearly indicate that the social context of disagreement affects children's evidence evaluation. The question is why does it do so? From many perspectives, another agent's hypotheses, and social information in general, should have little or no bearing on inferences from experimental data. Why, then, did we find disagreement to be helpful? One possibility harkens back to the classic finding that young and early school-age children often fail to differentiate theory and evidence (D. Kuhn, 1989; Zimmerman, 2007). Children might be treating the agent's hypothesis as an additional source of evidence. This explanation would suggest that children's sorts should move in the direction of the agent's hypothesis. This is consistent with our finding that children made stronger inferences in response to disagreement when the evidence was disconfirming. However, children did not always respond to the agent's hypothesis. What someone else thought only mattered when the data were somewhat ambiguous. Thus, children were not simply confusing theory for evidence.

A second possibility involves the salience of alternative hypotheses. By endorsing a hypothesis different from the child's, the agent provoked the child to consider evidence for and against the two alternatives. When evidence disconfirmed the child's hypothesis, the explicit endorsement of two alternative hypotheses (by the child and the agent) made it easier for the child to switch. This finding is reminiscent of claims in the history of science literature: Scientists are more willing to accept counterevidence and abandon an old theory when there is a viable alternative available (T. S. Kuhn, 1962). Note, though, that in this study, the counterhypothesis was always made salient for children. When asking the child and Leo to select a hypothesis, the experimenter twice explicitly pointed out that there were two viable hypotheses. We hypothesize that the role of the hypothesis in motivating the agent's intervention is important. Future studies should attempt to distinguish the general salience of alternatives from the significance of evidence generated by an agent operating with a specific counterhypothesis.

Note that salience of alternatives is not a complete explanation of the results. Even when it was salient, children weighed the agent's counterhypothesis against the data. When the data directly supported the child's hypothesis, the agent's hypothesis was irrelevant. When the child's hypothesis was undermined, having the alternative helped with revision. Older children showed an even more sensitive balancing of hypothesis and data: Even when the data were ambiguous, older children showed sensitivity to alternative hypotheses. Again, it may be that older children are better able to represent gradations of evidence. Children in both age groups recognized that an agent's hypothesis did not have the status of fact (unlike the result of the detector). Older children may appreciate that an agent's hypothesis is not simply making an alternative salient but rather may be providing some limited evidential support for that alternative. Christensen (2009) noted that, normatively, disagreement might be reason enough to revise one's beliefs only among peers with equally well-reasoned and informed beliefs. According to this view, older children in our experiment treated Leo as a well-reasoned peer more so than did the younger children. This suggests that the older children more readily imputed a rational basis to Leo's hypotheses, at least under circumstances of complete ambiguity.

A third potential explanation for the effects of disagreement holds that children reason about the outcomes of others' interventions in terms of the mental states that guide them. Because interventions are intentional actions, their informativeness may depend on the reasons why agents produced them (Kushnir et al., 2009). Identical interventions conducted under different hypotheses could be construed quite differently, rendering outcomes from dissenters' and nondissenters' interventions differentially informative. Studies of adult hypothesis testing and inductive reasoning confirm that contexts of disagreement and conflict can dramatically alter the design and interpretation of experiments (e.g., Butera, Caverni, & Rossi, 2005; Butera & Mugny, 2001). Perhaps children's better learning from disconfirmation when Leo disagreed is reflective of their imputing a rational basis for Leo's hypothesis and therefore giving heightened attention to outcomes of the subsequent interventions. Further research is needed to more fully situate these results in relation to a social cognitive account of interventionist causal learning.

Finally, a fourth potential explanation is that disagreement may facilitate attention to more aspects of a problem space. Many past findings are compatible with such a view. Although past studies have not focused on disagreement per se, they have focused on related constructs, such as sociocognitive conflict, contrasting cases, or counterexamples (e.g., Butera & Mugny, 2001; Karmiloff-Smith & Inhelder, 1974–1975; Namy & Gentner, 2002; Schwartz & Bransford, 1998). Peer disagreement and sociocognitive conflict generally seem to support more advanced solutions and cognitive growth. For example, Azmitia and Montgomery (1993) found that conflicting interpretations among peers were associated with more advanced experimental design (i.e., conducting unconfounded interventions). Similarly, research on the conditions that support collaborative learning highlights the importance of differing preconceptions among children (e.g., Doise & Mugny, 1984; Doise, Mugny, & Perret-Clermont, 1975; Webb, 1989). Note, however, that much of this research also emphasizes the need for discussion and resolution of contrasting perspectives (e.g., Kruger, 1993; Mercer & Littleton, 2007). Children in our study clearly had no discussion with Leo; the collaborative exchange was instantiated as a simple statement of alternative belief and a subsequent intervention and outcome—the mere statement of contradictory hypotheses led children to better evaluate experimental evidence in our task. These results support Howe's (2009) suggestion that processes surrounding unresolved conflict itself, and not group discussion/resolution of conflicting beliefs, account for collaborative learning gains in many cases.

It is important to note that the current study intentionally constrained the social context and task in several ways in order to directly investigate the effects of agent hypothesis. First and foremost, neither the children nor the agent had strong prior beliefs or expertise to motivate their hypotheses. Our approach is similar to that in the large literatures investigating children's basic evidence evaluation tendencies and causal reasoning skills in knowledge-lean settings with arbitrary and uniformly plausible hypotheses (see Zimmerman, 2007). Still, although children had little a priori reason to adopt one hypothesis over the other, their behavior reflected some degree of commitment to the hypotheses they endorsed. First, children sorted the $A+B-$ and $A-B+$ objects in accordance with their hypotheses before observing the relevant evidence. Second, children's final sorting of these objects showed differential effects of confirming and disconfirming evidence. Had children not committed to their hypotheses, the distinction between confirming and disconfirming evidence should have been null. Thus, at a minimum, our task successfully fostered belief in relatively arbitrary hypotheses and revealed how subsequent causal reasoning in response to various types of evidence can be affected via statements of disagreement concerning similarly plausible alternative hypotheses.

It is important to consider the extent to which disagreement in the experimental context we created resembles the kind of disagreement one might find in the classroom or science laboratory. Although our data suggest children did believe to a nontrivial degree in their endorsed hypotheses, such arbitrary hypothesis selection and simplicity is not a common occurrence in real-world settings. Both child learners and scientists often have strong theoretical commitments and experiences with relevant data that guide their hypothesis selection (Koslowski, 1996). Furthermore, scientific disagreement is driven by arguments concerning the differential plausibility of complex causal mechanisms over extended periods of time in rich social, institutional, and competitive contexts (Thagard, 1999). Our experimental setup clearly did not attempt to capture such aspects of authentic scientific disagreement. In future work, manipulating such factors should further illuminate the processes by which disagreement can affect evidence evaluation. Specifically, both the plausibility and strength of children's initial hypotheses could be manipulated using knowledge-rich domains or suggestive initial evidence. In addition, factors influencing children's evaluations of others' beliefs could be manipulated, including relative expertise and past reliability (Corriveau, Harris, et al, 2009; Harris & Koenig, 2006; Sobel & Corriveau, 2010). Similarly, the extent of disagreement could be manipulated by placing stakes on being correct (e.g., a competitive context) or by establishing a stronger history/pattern of dissent between the child and the agent. Finally, investigating disagreement in contexts of greater uncertainty (e.g., more complex causal systems or probabilistic evidence) would further align future work with real-world learning. In this study, we demonstrated significant effects of disagreement on learning in a simple, stripped-down context. This is an important first step and strongly suggests that more authentic forms of disagreement could dramatically impact reasoning in more complex, real-world situations.

An additional and essential extension of this work should examine disagreement among true peers (e.g., child dyads). Although using a naïve puppet allowed us to limit the inherent weight of the agent's agreement and disagreement, this choice was not without its limitations. Of course, interaction with a puppet fails to capture the richness of the social dynamics involved in disagreement between peers, and thus investigating disagreement between truly social partners will be necessary. Beyond improving ecological validity and generalizability, the use of a child peer or adult confederate could further illuminate the processes through which disagreement affects evidence evaluation. For example, earlier we suggested older children's response to disagreement in the neutral evidence condition might be explained by a greater tendency to treat Leo as an epistemic peer. Older children might also be more likely to treat Leo as an extension of the adult experimenter, potentially motivating the inference that Leo's beliefs have a rational basis. Direct manipulation of the type of peer (e.g., puppet, child, adult) could help distinguish these possible interpretations.

In addition to its theoretical contribution, this line of research could have future educational implications. Differences of opinion are common occurrences among school-age children and are seldom resolved (Hartup, French, Laursen, Johnston, & Ogawaw, 1993; Howe & McWilliam, 2001, 2006). Furthermore, social contextual factors, such as peer disagreement, are both ubiquitous in classrooms and highly malleable. As others (e.g., Howe, 2009) have suggested, facilitating or staging disagreement might be particularly fruitful for peers working on data evaluation (e.g., in a classroom science lab). Of course, the design of the present task (e.g., use of a puppet, simple causal systems, and generic hypotheses) does not yield direct and specific educational prescriptions. However, extensions of this paradigm (such as those mentioned above) could prove informative about children's group performance in the design of experiments (Azmitia & Montgomery, 1993), interpretation of experimental error and variability (Masnick, Klahr, & Morris, 2007), and general evidence evaluation. Understanding and identifying social contextual factors, such as disagreement, that influence children's evidence evaluation and causal reasoning will likely provide new insights into children's engagement in the practices of science and the optimal structuring of effective learning environments (National Research Council, 2007; Sobel & Sommerville, 2009).

To conclude, our results add to the growing literature that children's learning is affected in important ways by the social and pragmatic contexts in which data are generated. This study demonstrated that another's causal hypotheses can alter children's evaluations of evidence and subsequent beliefs about causality. Critically, children were quite successful in using the data to inform their causal judgments. Disagreement only mattered in situations of ambiguity. Thus, children used the social context to constrain their causal inferences in a sensible and productive way. There is often substantial ambiguity in the evidence supporting scientific and everyday causal beliefs, both inside and outside the lab. Perhaps we should expect consideration of the social contextual factors involved in evidence generation to be the norm rather than the exception.

References

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology, 18,* 894–897. doi:10.1037/0012-1649.18.6.894

Asch, S. E. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. *Psychological Monographs, 70*(9, Whole No. 416). doi:10.1037/h0093718

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues and development of scientific reasoning. *Social Development, 2,* 202–221. doi:10.1111/j.1467-9507.1993.tb00014.x

Butera, F., Caverni, J. P., & Rossi, S. (2005). Interaction with a high-versus low-competence influence source in inductive reasoning. *Journal of Social Psychology, 145,* 173–190. doi:10.3200/SOCP.145.2.173-190

Butera, F., & Mugny, G. (2001). Conflicts and social influences in hypothesis testing. In C. K. W. De Drue & N. K. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation* (pp. 160–182). Cambridge, MA: Blackwell.

Chandler, M., & Lalonde, C. (1996). Shifting to an interpretive theory of mind: 5- to 7-year-olds' changing conceptions of mental life. In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 111–139). Chicago, IL: University of Chicago Press.

Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass, 4,* 756–767. doi:10.1111/j.1747-9991.2009.00237.x

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer non-dissenters as informants. *Psychological Science, 20,* 372–377. doi:10.1111/j.1467-9280.2009.02291.x

Corriveau, K. H., Harris, P. L., Meins, E., Ferneyhough, C., Arnott, B., Elliott, L., . . . de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80,* 750–761. doi:10.1111/j.1467-8624.2009.01295.x

Doise, W., & Mugny, G. (1984). *The social development of the intellect.* Oxford, England: Pergamon.

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5,* 367–383. doi:10.1002/ejsp.2420050309

Feist, G. J. (2006). *The psychology of science and the origins of the scientific mind.* New Haven, CT: Yale University Press.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60,* 115–140. doi:10.1146/annurev.psych.59.103006.093659

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,* 3–32. doi:10.1037/0033-295X.111.1.3

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

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145–171. doi:10.1111/j.1468-0017.1992.tb00202.x

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

Harris, P. L., & Koenig, M. (2006). Trust in testimony: How children learn about science and religion. *Child Development, 77,* 505–524. doi:10.1111/j.1467-8624.2006.00886.x

Hartup, W. W., French, D. C., Laursen, B., Johnston, M. K., & Ogawa, J. R. (1993). Conflict and friendship relations in middle childhood: Behavior in a closed-field situation. *Child Development, 64,* 445–454. doi:10.2307/1131261

Howe, C. (2009). Collaborative group work in middle childhood: Joint construction, unresolved contradiction and the growth of knowledge. *Human Development, 39,* 71–94.

Howe, C., & McWilliam, D. (2001). Peer argument in educational settings: Variation due to socioeconomic status, gender and activity context. *Journal of Language and Social Psychology, 20,* 61–80. doi:10.1177/0261927X01020001003

Howe, C., & McWilliam, D. (2006). Opposition in social interaction amongst children: Why intellectual benefits do not mean social costs. *Social Development, 15,* 205–231. doi:10.1046/j.1467-9507.2006.00337.x

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York, NY: Basic Books. doi:10.1037/10034-000

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology, 61,* 248–272. doi:10.1016/j.cogpsych.2010.06.002

Karmiloff-Smith, A., & Inhelder, B. (1974–1975). If you want to get ahead, get a theory. *Cognition, 3,* 195–212. doi:10.1016/0010-0277(74)90008-0

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes.* Cambridge, MA: MIT Press.

Koenig, M. A., & Harris, P. L. (2005). The role of social cognition in early trust. *Trends in Cognitive Sciences, 9,* 457–459. doi:10.1016/j.tics.2005.08.006

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

Kruger, A. C. (1993). Peer collaboration: Conflict, cooperation or both? *Social Development, 2,* 165–182. doi:10.1111/j.1467-9507.1993.tb00012.x

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689. doi:10.1037/0033-295X.96.4.674

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2008). The role of preschoolers' social understanding in evaluating the informativeness of causal interventions. *Cognition, 107,* 1084–1092. doi:10.1016/j.cognition.2007.10.004

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2009). A self-agency bias in children's causal inferences. *Developmental Psychology, 45,* 597–603. doi:10.1037/a0014727

Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology, 20,* 163–198. doi:10.1016/S0732-118X(02)00007-7

Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. Lovett & P. Shaw (Eds.), *Thinking with data* (pp. 3–26). Mahwah, NJ: Erlbaum.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking.* London, England: Routledge.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131,* 5–15. doi:10.1037/0096-3445.131.1.5

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8* (R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds.). Washington, DC: National Academies Press.

Oreskes, N. (1999). *The rejection of continental drift: Theory and method in American earth science.* New York, NY: Oxford University Press.

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis–evidence relation. *Child Development, 64,* 1617–1636. doi:10.2307/1131459

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102–119. doi:10.1037/0012-1649.32.1.102

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared cause tummy aches? Naïve theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43,* 1124–1139. doi:10.1037/0012-1649.43.5.1124

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522. doi:10.1207/s1532690xci1604_4

Schwarz, B., Neuman, Y., & Biezuner, S. (2000). Two wrongs make a right . . . If they argue together! *Cognition and Instruction, 18,* 461–494. doi:10.1207/S1532690XCI1804_2

Silverman, M. P. (1992). Raising questions: Philosophical significance of controversy in science. *Science & Education, 1,* 163–179. doi:10.1007/BF00572837

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81,* 669–679. doi:10.1111/j.1467-8624.2009.01422.x

Sobel, D. M., & Sommerville, J. A. (2009). Rationales in children's causal learning from others' actions. *Cognitive Development, 24,* 70–79. doi:10.1016/j.cogdev.2008.08.003

Sommerville, J. A., & Hammond, A. J. (2007). Treating another's actions as one's own: Children's memory of and learning from joint activity. *Developmental Psychology, 43,* 1003–1018. doi:10.1037/0012-1649.43.4.1003

Thagard, P. (1999). *How scientists explain disease.* Princeton, NJ: Princeton University Press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning, *Behavioral and Brain Sciences, 16,* 495–552. doi:10.1017/S0140525X0003123X

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21–39. doi:10.1016/0883-0355(89)90014-1

Williams, J., & Tolmie, A. (2000). Conceptual change in biology: Group interaction and the understanding of inheritance. *British Journal of Developmental Psychology, 18,* 625–649. doi:10.1348/026151000165896

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27,* 172–223. doi:10.1016/j.dr.2006.12.001

Appendix

Object Set Features

Set	Feature 1	Feature 2
Practice	Red, blue	Cube, sphere
Set A	Large, small	Spikes, loops
Set B	Furry, smooth	One stick, two sticks
Set C	Tall, short	Holes, no holes
Set D	T-shaped, X-shaped	Spring, knob
Set E	Green, purple	Head, tail
Set F	Few stripes, many stripes	Solid, stretchy

Received June 26, 2010
Revision received December 15, 2011
Accepted January 9, 2012 ■