CHILD DEVELOPMENT

Child Development, May/June 2019, Volume 90, Number 3, Pages 859–875

Causal Learning Across Culture and Socioeconomic Status

Adrienne O. Wente and Katherine Kimura
University of California, Berkeley

Caren M. Walker
University of California, San Diego

Nirajana Banerjee
University of California, Berkeley

María Fernández Flecha
Pontificia Universidad Católica del Perú

Bridget MacDonald
University of California, San Diego

Christopher Lucas
The University of Edinburgh

Alison Gopnik
University of California, Berkeley

Extensive research has explored the ability of young children to learn about the causal structure of the world from patterns of evidence. These studies, however, have been conducted with middle-class samples from North America and Europe. In the present study, low-income Peruvian 4- and 5-year-olds and adults, low-income U.S. 4- and 5-year-olds in Head Start programs, and middle-class children from the United States participated in a causal learning task (N = 435). Consistent with previous studies, children learned both specific causal relations and more abstract causal principles across culture and socioeconomic status (SES). The Peruvian children and adults generally performed like middle-class U.S. children and adults, but the low-SES U.S. children showed some differences.

Coming to understand the causal structure of the world is a central part of cognitive and conceptual development. Causal learning plays an especially important role in the development of intuitive theories of the world, such as folk biology and "theory of mind." In the past 15 years, there has been a large body of research showing how computational systems can accurately infer causal relations from statistical patterns of data (e.g., Pearl, 2009; Spirtes, Glymour, & Scheines, 1993/2000; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). There have also been hundreds of causal learning experiments with toddlers and preschoolers, coming from a number of different laboratories (see Gopnik & Wellman, 2012; Xu & Kushnir, 2012 for recent reviews). These studies show remarkably high levels of competence in young children. In fact, some recent causal learning studies have yielded a counterintuitive pattern of findings: The ability to infer certain types of abstract causal relations from evidence actually appears to decline with age, so that younger children do better than older children and adults (Gopnik, Griffiths, & Lucas, 2015; Gopnik et al., in press; Lucas, Bridgers, Griffiths, & Gopnik, 2014; Seiver, Gopnik, & Goodman, 2013).

However, to our knowledge, these studies have all examined children in similar middle- to upper-middle-class samples in North America and Europe—that is, children from Western, educated, industrial, rich, and democratic (WEIRD) cultures (Henrich, Heine, & Norenzayan, 2010). There has been important cross-cultural work on the development of intuitive theories (e.g., Avis & Harris,

Adrienne O. Wente and Katherine Kimura contributed equally to this work.

This research was supported by funding from the Bezos Family Foundation and Carlos Rodriguez-Pastor. The authors would like to thank Alexa Kossuth Dasso and Maria Fe Hashimoto for data collection in Peru. They also wish to thank Teresa Garcia, Renata Botello, and Aaya Abdelshafy for assistance with data collection and coding in the United States, as well as the parents, students, teachers, and administrators at the Innova Schools in Peru, Berkeley and Oakland Head Start Schools, the Early Childhood Education Program at UC Berkeley, the Berkeley School Early Childhood Center, Montverde School, and Step One School.

Correspondence concerning this article should be addressed to Adrienne O. Wente, Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720. Electronic mail may be sent to adrieorchid@berkeley.edu.

 0009-3920/2019/9003-0013
DOI: 10.1111/cdev.12943

1991; Callaghan et al., 2005; Coley, 2012; Gelman & Legare, 2011; Medin & Atran, 2004; Wellman, Fang, Liu, Zhu, & Liu, 2006), but not on the causal learning abilities that might help underpin this development.

In this article, we extend research on causal learning to two very different types of populations: a cross-cultural sample of relatively low-income Peruvian children and adults and a cross-socioeconomic sample of children from low-income North American families. There is increasing recognition of the importance of including a wider range of participants in developmental studies (e.g., Henrich et al., 2010; Legare & Harris, 2016). A recent review article, for example, found that < 1% of published developmental studies were conducted in South or Central American countries, with important implications for the generality of the conclusions drawn from those studies (Nielsen, Haun, Kärtner, & Legare, 2017). Studying causal learning, in particular, across these populations may be especially important from a practical perspective. Causal learning is itself a central part of cognitive development and is important for the development of intuitive theories. The ability to infer causal structure from evidence is also a foundational part of scientific reasoning. Differences in these abilities, then, might have important consequences for education in general and science education in particular.

Previous Studies

Most of the studies of causal learning in children have focused on children's ability to learn particular causal hypotheses from statistical patterns of data, particularly covariation between causes and effects. Early studies found that children as young as 24 months old could make these inferences in ways that went beyond simple associative learning (Gopnik, Sobel, Schulz, & Glymour, 2001; Gopnik et al., 2004; Sobel & Kirkham, 2007). Researchers also found that children could use statistical patterns of evidence to infer more complex structures, such as common causes versus causal chains (Schulz, Gopnik, & Glymour, 2007) and to infer unobserved causes (Gopnik et al., 2004; Schulz & Sommerville, 2006) and that children as young as 24 months old could infer probabilistic as well as deterministic causal relations (Kushnir & Gopnik, 2005; Waismeyer, Meltzoff, & Gopnik, 2014).

More recently, researchers have investigated whether children can also learn more abstract and general causal principles, the equivalent of "framework theories" (Gopnik & Wellman, 2012; Laudan, 1978) or "overhypotheses" (Goodman, 1955) in science. These principles do not specify particular causal hypotheses, but they do constrain the possible hypotheses a learner will consider (Goodman, 1955; Griffiths & Tenenbaum, 2007). For example, in intuitive psychology an "overhypothesis" would specify that the causal explanation of an action is likely to involve beliefs and desires—even if it does not specify exactly which beliefs and desires explain a particular action. Several studies from different laboratories show that preschoolers can also make these more abstract inferences (Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Seiver et al., 2013; Sim & Xu, 2017).

In particular, in three separate studies, Lucas et al. (2014) contrasted two abstract overhypotheses: one that is initially more likely and one that is less likely (at least for adults). Participants received evidence of the covariation between causes and effects that supported the more likely overhypothesis or else received covariation evidence that supported the less likely overhypothesis. Experimenters then presented the participants with a new set of evidence and recorded whether participants of different ages interpreted that evidence in accord with the overhypothesis. In more detail, preschoolers and adults saw evidence in a training trial that justified the conclusion that a novel machine worked according to one of two abstract logical principles: a disjunctive (i.e., OR) principle or a conjunctive (i.e., AND) principle. Almost all the earlier studies of children tested disjunctive causal relations, which are also often the default assumption in studies of adults (see Cheng, 1997). However, one study (Schulz et al., 2007) found that 4-year-olds would infer a conjunctive causal structure given the right evidence.

Participants were first told that "blicketness" activated the machine. If participants were assigned to the disjunctive condition, they saw that a single object would or would not activate the machine. Those assigned to the conjunctive condition, however, observed that only a combination of two objects would activate the machine. After viewing this evidence, participants saw an ambiguous activation pattern with new objects that was consistent with both the disjunctive and conjunctive principles, and had to infer the causal structure underlying that pattern (see Figure 1).

Participants demonstrated their causal inferences in two ways. First, they judged whether each of three objects, D, E, or F, was or was not a blicket. Second, they intervened to make the machine go.

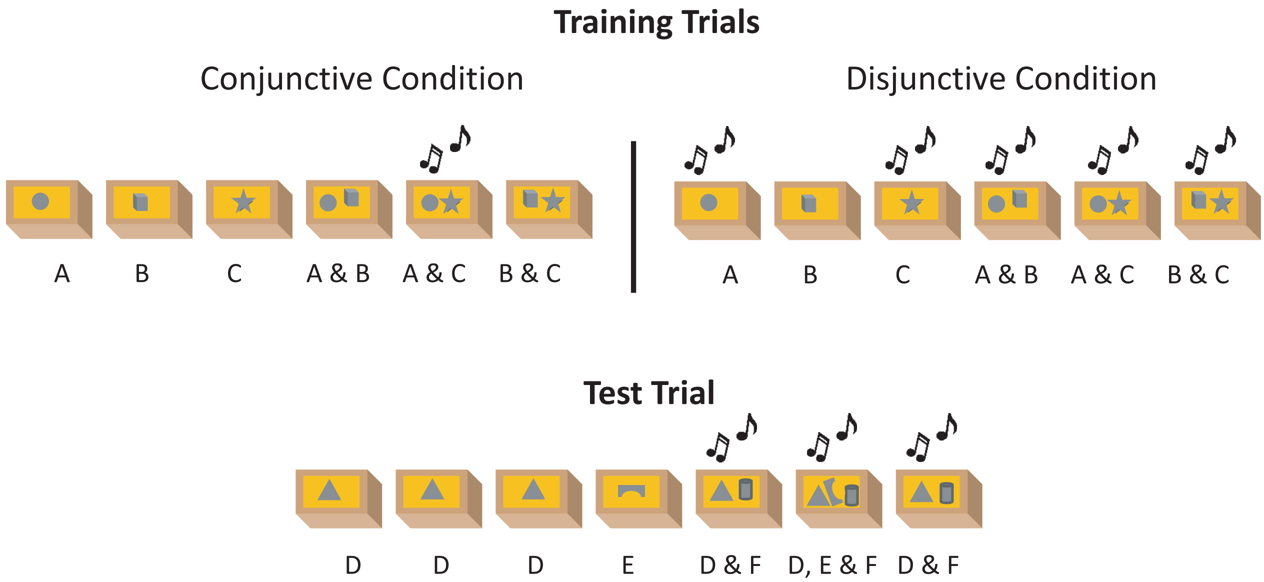

Figure 1. Activation patterns in the conjunctive and disjunctive conditions and test trial.

Given the ambiguous evidence, if they had inferred the disjunctive principle, they should say that only F was a blicket and should use just that object to activate the machine. If they had inferred the conjunctive principle, they should say that more than one object was a blicket—in particular, both D and F would be more likely to be blickets than E—and they should choose multiple objects to activate the machine.

Importantly, the disjunctive and conjunctive conditions were completely identical except for the activations in the training trials. The same objects were presented in the same order with the same script in the training trials, and the participants made judgments about exactly the same ambiguous test trial. The conditions could therefore act as controls for each other. If participants responded systematically differently in the two conditions, this suggests that they were influenced by the training pattern. Similarly, comparing the judgments of the three objects across the two conditions ensures that participants do not simply have a tendency to say that objects are "blickets" (or not), or to behave at chance, but are paying attention to the specific pattern of evidence in each condition. In addition, in the disjunctive training trials, participants actually see more examples of blocks activating the machine, and more multiple combinations of blocks activating the machine, than they do in the conjunctive trials, where the machine only activates once (see Figure 1). Nevertheless, the correct causal inference involves the opposite response, saying that more blocks are blickets and placing more objects on the machine in the conjunctive than disjunctive condition. If children produce these responses, it suggests that they are not using a simple imitative or perceptual strategy.

Lucas et al. (2014) found that middle-class U.S. preschool children readily learned both types of abstract causal rules depending on the activation pattern they observed. This means that, given the general causal principle that the machine was disjunctive or conjunctive, children inferred the correct specific causal hypothesis in the test phase, concluding exactly which blocks were most likely to activate the machine. But significantly, it also means that they were inferring the general principle itself from the training data, and then applying that general principle to the ambiguous test data. So the task both tests children's abilities to infer specific causal hypotheses and more general overhypotheses from data. In the current article, we extend this method across cultures and socioeconomic status (SES).

Notably also, adults, unlike the preschoolers, tended to infer that the machine was activated by one object, thus following the more likely disjunctive principle regardless of the training evidence. A recent study (Gopnik et al., 2017) further tracked this ability across development. Six- to 11-year-olds were less likely to infer the unusual conjunctive structure than 4-year-olds, and 12- to 14-year-olds as well as adults were less likely to infer the conjunctive structure than school-aged children.

Research Questions and Predictions

Cross-cultural and cross-SES studies are intrinsically and practically important, but they can also help to differentiate among different, more specific, hypotheses about development. First, it is possible that the precocious causal learning seen in earlier studies depends on growing up in an enriched environment with many artifacts and extensive informal science-like pedagogy. Since the ability to infer causal structure from evidence is a foundational part of scientific methodology, it might be encouraged in cultures or households where science is valued. Indeed, many studies have recruited children from science museums and from university preschools, and it is plausible that these parents do provide implicit science pedagogy. Moreover, the information-processing demands of tasks such as that in Lucas et al. (2014) are quite high—involving tracking nine objects and 25 activation events. Children in other cultures or classes might prove to have more information-processing difficulty than the WEIRD children in previous studies. In either case, we might expect to see general difficulties and lower performance on all these tasks, involving both the usual and unusual "overhypotheses," in children from other backgrounds.

For similar reasons, children in other cultural groups might demonstrate some causal learning abilities but not others, suggesting a developmental ordering. In particular, children might be able to correctly infer the more common disjunctive structure, but have difficulty with the less common conjunctive one. They also might simply default to one hypothesis or the other in the ambiguous test trials, but fail to differentiate between them, suggesting that they had not inferred the overhypothesis from the training trials.

Alternatively, children (and adults) might be able to solve these tasks in general, but they might show different patterns of response across cultures. For example, different cultural groups might have different preferred "overhypotheses" about causal relations. In particular, cross-cultural psychologists frequently distinguish between analytic and holistic styles of causal reasoning (Henrich et al., 2010; Nisbett, Peng, Choi, & Norenzayan, 2001; Peng & Knowles, 2003). Individuals from Western cultures, for example, more readily engage in an analytic style of reason, which involves identifying a single candidate cause. In contrast, the holistic style of reasoning places a greater emphasis on the causal relations between multiple objects, and between objects and their environment. Individuals from analytic backgrounds may learn to attribute causation to a primary, focal object, whereas those who reason more holistically may more readily assume a cause that requires a combination of objects. There is research suggesting that people from some South American cultures reason more holistically than people from the United States (Henrich et al., 2010). In fact, Henrich et al. (2010) argue that mainstream U.S. culture falls on the extreme end of the distribution and that people from the United States reason more analytically than people from most other cultures. Given this, another possibility is that children and adults from Peru, particularly from groups that have recently migrated from rural indigenous communities, might actually be more likely to infer the multiple object conjunctive structure than children and adults from the United States.

Finally, the pattern of results and levels of performance might be similar across SES and culture, suggesting that these abilities are more generally characteristic of young children.

Experiment 1

Method

Participants

Ninety Peruvian 4- and 5-year-old children (45 females and 45 males), and 145 Peruvian undergraduate college students (88 females and 57 males) participated in this study. Three children were tested and not included in the analysis. One child was excluded because of experimenter error and two children were excluded because they picked up and placed one of the objects on or near the machine at an inappropriate time. No adults were excluded.

Thirty children participated in the conjunctive condition (M = 5.21, range = 4.21–5.97), 30 participated in the disjunctive condition (M = 5.18, range = 4.22–6.01), and 30 participated in the baseline condition (M = 5.2, range = 4.09–6). Forty-nine Peruvian adults participated in the conjunctive condition (M = 22.72, range = 19.97–39.76), 44 participated in the disjunctive condition (M = 25.92, range = 16.93–45.15), and 52 participated in the baseline condition (M = 19.45, range = 17.11–24.63). The published data set from Lucas et al. (2014) was used to compare results across cultures.

Children were recruited and tested at Innova schools in Lima, Peru. Innova schools are a Peruvian chain of private schools designed to provide affordable education to Peru's emerging middle class (e.g., families who are transitioning from low-income to middle-income). In some ways, this sample was similar to the children in Lucas et al. (2014). Children from both backgrounds were from urban, industrialized environments, and parents were heavily invested in their education. However, the samples also differed along key dimensions. Not only were children tested in different countries, but Peruvian children also had a substantially lower SES than the Lucas et al. (2014) children. Innova schools were frequently located in low-SES, high crime neighborhoods within Lima or in less developed neighborhoods on the outskirts of Lima, accessible only by dirt roads.

In general, the average income of families across all the Innova schools is around $1,200 per month. However, we chose to test in schools at the lower end of the income spectrum, so the children were from families who almost certainly made less than $1,200 per month on average. Only a small percentage of these parents have had access to higher education and most are small business owners. Spanish was the native language of the children, but most students were second- or third-generation internal immigrants from the Andean highlands with a strong indigenous heritage.

Peruvian undergraduate students were recruited and tested at Pontificia Universidad Católica del Perú, a local private university often chosen by the emerging middle class. Although the tuition varies according to the family's income—it can vary between $333 and $1,363 a month—almost 70% of students are among those who pay lower tuition. Many of the students are also first-generation college students.

Materials

A machine derived from Gopnik and Sobel (2000) was constructed from a wooden box that was approximately 10 × 6 × 5 in. The top of the box was covered with orange construction paper and a wireless doorbell was placed inside. This doorbell was surreptitiously activated by the experimenter. Fifteen 3-dimensional clay blocks of unique geometric shapes but of similar sizes were painted gray for color uniformity. These objects resembled those used by Lucas et al. (2014). While the exact dimensions varied, objects were about 3 in. and were made into a variety of distinctive shapes, such as a square, a circle, and a star. Three plastic semiopaque buckets held the objects. Buckets were approximately 7 × 5 × 3 in.

Procedure

The method was identical to Lucas et al. (2014) Experiment 2. Children were tested one by one in a private office or classroom in their school. Adults were tested in university classrooms in groups of approximately 20 subjects per testing session, similar to the procedure in Lucas et al. (2014). An experimenter demonstrated the procedure, and adults were instructed to write their answers in booklets so as to avoid any contamination of responses.

The word "flipo" was used as a plausible nonsense word in Spanish rather than "blicket." The experimenter first introduced participants to a bucket of nine unsorted objects and the machine. They explained that flipos contained a property called "fliponess" which activated the machine. They also explained that flipos were quite rare, and that the goal of the game was to figure out which of the objects were flipos. Next, the experimenter presented two new buckets, said that someone else had already tested these objects on the machine, and showed that one bucket contained a single flipo, while the other bucket contained four nonflipos.

Following this, the experimenter selected three objects, in what appeared to be a random order, from the first bucket, which contained the nine unsorted objects, and asked the child to name the objects. The order in which objects were selected was counterbalanced. Children and adults then received two training trials, with objects A, B, and C, and A1, B1, and C1. Each training event varied according to condition (see Figure 1), and was consistent with one of the two overhypotheses, conjunctive or disjunctive. In particular, in the conjunctive training, a combination of objects A and C was necessary to activate the machine. In the disjunctive training, either A or C activated the machine. Each trial also involved different objects. After observing each trial, the experimenter asked participants whether each object was a flipo. For example, the experiment might say, "Do you think the Star is a flipo or not a flipo?" No feedback was provided.

After the two training events, the experimenter brought out another object (G) which she said she had forgotten. This object was never placed on the machine, and children and adults were asked to guess if they thought the object was a flipo.

Finally, the experimenter selected three more objects: D, E, and F. The experimenter then demonstrated a new and ambiguous event (see Figure 1). During this event, it was not clear if one object alone activated the machine or if a combination of objects activated the machine. After demonstrating the event, the experimenter once again asked the participant to say whether each object was a flipo or not a flipo. Afterward, children and adults were given an intervention prompt, and were asked about the object(s) they would use to activate the machine.

The procedure for the baseline condition was similar to the disjunctive and conjunctive experimental conditions except that it omitted the two training events. In the baseline, participants also provided feedback for two test events rather than one. We analyzed and reported their responses to the first test event.

Translation

The study protocol from Lucas et al. (2014) was translated to Spanish by an experienced Peruvian linguist as well as back-translated to English in order to assess the translation. The English version of this question, "Do you think {insert shape name} is a Blicket or not a Blicket?" was directly translated to, "¿Crees que {insert shape name} es un flipo o no es un flipo?"

The second question was the intervention prompt. In English it read, "Which of these should we use to make my machine turn on?" Translating this directly to Spanish was challenging since the English word, "which," can be translated into either "¿Cuál?" or "¿Cuáles?" These, however, provide a singular–plural marking, which could indicate to participants that they should select either one object or multiple objects for the intervention. Selection of single versus multiple objects is a dependent variable. To circumvent this, the intervention prompt for the children was translated to, "¿Qué harías para que la máquina se encienda?" which directly translates in English to, "What would you do to make the machine turn on?" This prompt avoids the use of the singular or plural marking; however, it also makes it much less clear that the intervention refers specifically to choosing among the three objects on the table. In fact, the most effective technique for activating the machine in all conditions would be to place all the objects on the machine since that would guarantee that the machine would light up.

An initial inspection of children's responses suggested that this difference in language did indeed lead to different intervention patterns. There were no significant differences in the intervention responses across conditions for Peruvian children, and they tended to select more objects than American children overall. Because of this concern, the intervention prompt was retranslated for adults to, "¿Qué deberíamos usar de aquí para hacer que la máquina se encienda?" This translates in English to, "What should we use from here to make the machine turn on?" This question still excluded the singular–plural marking yet was somewhat more comparable to the original English prompt in that it explicitly guided participants to select from the three objects directly displayed in front of them. However, it still did not make specific reference to the blocks and so might have led to more variable responses. This second prompt was used when testing adults.

Inspection of adult data indicates that Peruvian adults, like the Peruvian children, tended to choose multiple objects for interventions more than U.S. adults: 44% of the Peruvian adults chose multiple objects, whereas only 23% of the U.S. adults did so, although their flipo judgments were comparable. One reason for this might be that the English word "which" encouraged English-speaking participants to be more discriminatory in their selection, whereas the Spanish phrasing using "que" or "what" is more suggestive of multiple items. Because the questions were not comparable across the languages and across children and adults, we did not include further analysis of the intervention data in the Results section of Experiment 1.

Results

Training Trials: A Versus B Versus C

We first assessed participants' responses during the training trials. This served to test whether children could infer specific causal hypotheses from unambiguous evidence, and whether they attended to and understood those trials. In the training trials, participants in the disjunctive condition should conclude that both A and C activate the machine, although they do so independent of each other, but that B does not. Those in the conjunctive condition should, likewise, infer that objects A and C are flipos, because both are necessary together to activate the machine, and that B is not. (Note that the actual A, B, and C objects shapes differed across participants and training trials, and were counterbalanced; see Figure 1.)

If participants are confused by the task or have difficulty processing the activation patterns, they should fail to differentiate between objects A, B, and C, and instead respond "yes" (or "no") to all three objects or respond at chance when asked whether the blocks were flipos. Figure 2 illustrates the proportion of children and adults who judged A, B, and C as flipos.

Overall, we found that participants in both conditions successfully differentiated object B from objects A and C and thus were not simply responding "yes," to each question. Participants received a score of 0, 1, or 2 depending on how often they said each block was a blicket across the two trials. As predicted, children and adults in the *disjunctive* condition reliably judged object A (children: $M = 1.57$, $SD = 0.82$; adults: $M = 1.86$, $SD = 0.41$) to be a flipo more often than they judged object B to be a flipo (children: $M = 0.30$, $SD = 0.65$; adults: $M = 0.18$, $SD = 0.50$), $t(58) = 6.639$, $p < .001$; $t(86) = 17.375$, $p < .001$, respectively. They were also more likely to call object C a flipo (children: $M = 1.43$, $SD = 0.77$; adults: $M = 1.86$, $SD = 0.46$) than object B, $t(58) = 6.137$, $p < .001$; $t(86) = 16.47$, $p < .001$, respectively.

Similar patterns were observed in the *conjunctive* condition, with children and adults once again reporting that objects A (children: $M = 1.57$, $SD = 0.73$; adults: $M = 0.57$, $SD = 0.79$) and C (children: $M = 1.43$, $SD = 0.73$; adults: $M = 0.39$, $SD = 0.70$) were flipos more often than B (children: $M = 0.37$, $SD = 0.62$; adults: $M = 0.06$, $SD = 0.24$), A versus B—children: $t(58) = 6.897$, $p < .001$; adults: $t(56.93) = 4.319$, $p < .001$; B versus C—children: $t(58) = 6.131$, $p < .001$; adults: $t(59.28) = 3.079$, $p = .003$, respectively.

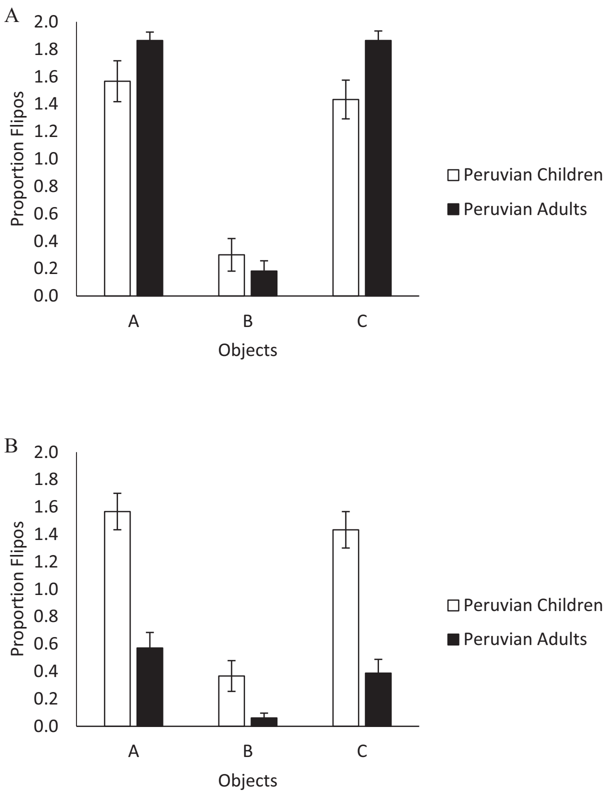

Figure 2. Average proportion of flipo judgments for objects A, B, and C among Peruvian children and adults in the disjunctive condition (A) and conjunctive condition (B). Error bars indicate 1 *SEM* in each direction.

Test Trials

D, E, and F flipo judgment comparisons. Participants were clearly differentiating the objects in the training trials; we next explored whether they were also differentiating objects in the test trial. If participants reason conjunctively, then they should infer that objects D and F are flipos—since together they activate the machine three times—but they should be uncertain about object E, which is associated with one conjunctive activation and one nonactivation, even though E and D are equally strongly associated with the activation of the machine (see Lucas et al., 2014 for discussion). If they reason disjunctively, then they should infer that object F is a flipo, whereas objects D and E are not. If, however, participants are confused by the task, they should respond "yes" or "no" to all three objects or respond at chance. The crucial contrasts, then are between D and F, which should be different in the disjunctive condition and similar in the conjunctive one, and between D and E which should be similar in the disjunctive condition and different in the conjunctive one (see Figure 3 for participants' D, E, and F judgments).

In the *disjunctive condition*, Peruvian children called object F a flipo ($M = 0.77$, $SD = 0.43$) reliably more often than object D ($M = 0.27$, $SD = 0.45$), $p < .001$, McNemar's exact test, or object E ($M = 0.20$, $SD = 0.41$), $p < .001$, McNemar's exact test. There were no significant differences between objects D and E, $p = .683$, McNemar's exact test. This is the pattern we would expect if learners were making the correct inference from the evidence.

In the *conjunctive condition*, children reliably reported that D was a flipo ($M = 0.87$, $SD = 0.35$) more often than E ($M = 0.63$, $SD = 0.49$), $p = .046$, McNemar's exact test. There was no significant difference between performance on D and F, $p = .683$, McNemar's exact test. Again, this is the pattern of responses we would expect if participants learned the correct causal structure. Therefore, despite this challenging and unusual task, Peruvian children were successfully tracking the activation patterns to

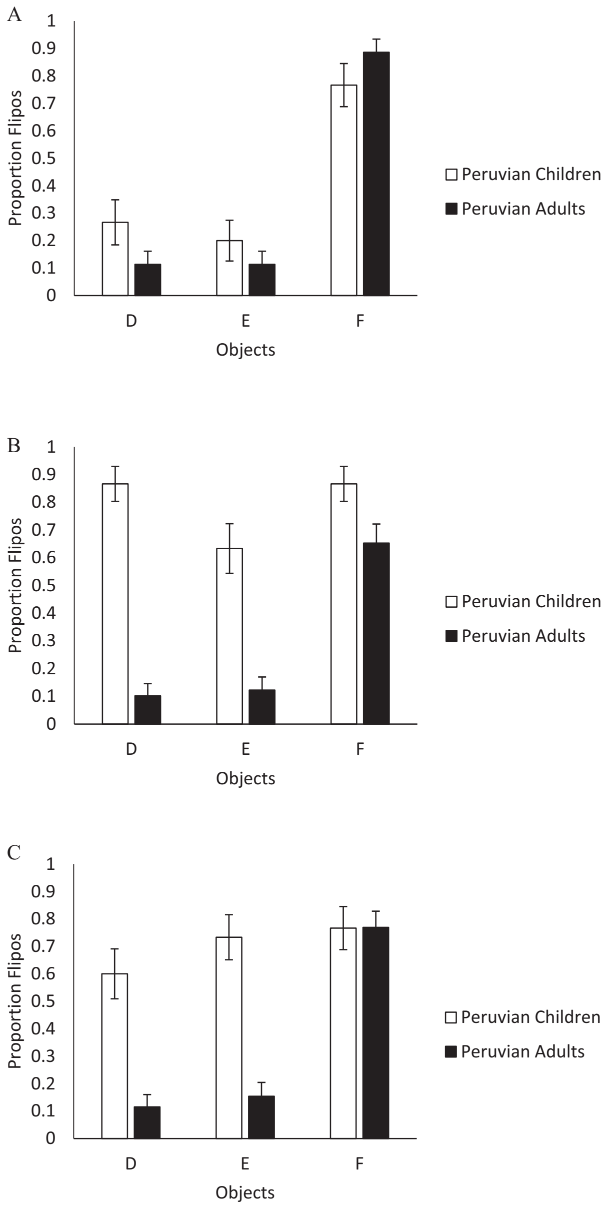

Figure 3. Average proportion of flipo judgments for objects D, E, and F among Peruvian children and adults in the disjunctive condition (A), conjunctive condition (B), and baseline condition (C). Error bars indicate 1 *SEM* in each direction.

make accurate causal inferences and did not appear to be confused.

In the *baseline condition*, for which children saw neither evidence supporting the conjunctive nor evidence supporting the disjunctive principle, children were just as likely to label object F (M = 0.77, SD = 0.43) a flipo as they were to label objects D (M = 0.60, SD = 0.50), p = .131, McNemar's exact test, and E (M = 0.73, SD = 0.45), p = 1.0, McNemar's exact test.

McNemar's exact tests were also used to analyze adult's responses. In the *disjunctive condition*, Peruvian adults called object F a flipo (M = 0.89, SD = 0.32) more often than object D (M = 0.11, SD = 0.32), $p < .001$. There was no significant difference between adults' judgments of objects D and E (M = 0.11, SD = 0.32), p = .480.

However, Peruvian adults in the *conjunctive condition*, as in the disjunctive condition, also called object F a flipo (M = 0.65, SD = 0.48) more often than object D (M = 0.10, SD = 0.31), $p < .001$. There was no difference in their judgments of objects D and E (M = 0.12, SD = 0.33), p = 1.0.

Similar patterns were observed in the *baseline condition*. Peruvian adults were more likely to say that object F was a flipo (M = 0.77, SD = 0.43) than object D (M = 0.12, SD = 0.32), $p < .001$. We found no reliable difference, however, in adults' judgments of objects D and E (M = 0.15, SD = 0.36), p = .752.

Flipo judgments of D and F across condition. The crucial responses to determine if participants inferred the overhypotheses were the judgments of D across the conditions. If participants concluded that the machine operated on a disjunctive principle, they should say that F was a flipo and D was not. If they concluded that the machine operated on a conjunctive principle, they should say both F and D were flipos.

As predicted, both age groups in all three conditions tended to say that F was a blicket (responses ranging from 65% to 89%, see Table 1). To test whether Peruvian children and adults differed in their judgments of D as a function of condition, we performed a binary logistic regression with age group, condition, and the interactions between each of these as predictors of judgments of D. The resulting model was statistically significant, $\chi^2(5) = 83.361$, $p < .001$, Nagelkerke $R^2 = .427$. The results also yielded a main effect of age group, $\chi^2 = 32.032$, $df = 1$, $p < .001$, and condition, $\chi^2 = 18.603$, $df = 2$, $p < .001$, and an interaction between the two, $\chi^2 = 9.961$, $df = 2$, $p = .007$. To more closely examine these effects, we next looked at performance across conditions in each age group using two-tailed Fisher's exact tests.

Peruvian Children

Crucially, Peruvian children in the conjunctive condition were significantly more likely than those in the disjunctive condition to call object D a flipo, $p < .001$, Fisher's exact tests. Children in the conjunctive condition also labeled object D a flipo more

Table 1
Mean Proportion of Flipo or Blicket Judgments for Peruvian Children and Adults in Experiment 1 and North American Children and Adults From Lucas et al. (2014)

	Flipo or blicket judgments		
Participant group	D	E	F
Peruvian children			
Disjunctive (*N* = 30)	.27	.20	.77
Conjunctive (*N* = 30)	.87	.63	.87
Baseline (*N* = 30)	.60	.73	.77
Peruvian adults			
Disjunctive (*N* = 44)	.11	.11	.89
Conjunctive (*N* = 49)	.10	.12	.65
Baseline (*N* = 52)	.12	.15	.77
North American children (Lucas et al., 2014; Experiment 2)			
Disjunctive (*N* = 25)	.32	.28	.80
Conjunctive (*N* = 25)	.92	.68	.88
Baseline (*N* = 24)	.42	.33	.75
North American adults (Lucas et al., 2014; Experiment 2)			
Disjunctive (*N* = 28)	.11	.11	.82
Conjunctive (*N* = 28)	.25	.11	.71
Baseline (*N* = 26)	.08	.08	.81

often than those in the baseline condition, $p = .039$, Fisher's exact test.

In general, the pattern of judgments was very similar to the pattern in Lucas et al. (2014). Children inferred the conjunctive hypothesis in the conjunctive case and vice versa, and their performance in the baseline condition appeared to fall in between. In fact, a direct comparison with the Lucas et al. results on every measure using two-tailed Fisher's exact tests found only one difference: The Peruvian children were more likely than the U.S. children to select E as a "flipo" in the baseline condition, $p = .006$ (see Table 1 for a cross-cultural comparison).

Peruvian Adults

In contrast, there were no significant differences among adults across conditions in their judgments of D (Fisher's exact tests on D judgments: conjunctive vs. disjunctive, $p = 1.00$; baseline vs. conjunctive, $p = 1.00$; baseline vs. disjunctive, $p = 1.00$). Fisher's exact tests also revealed no significant differences on any measure between the Peruvian adults and the adults in Lucas et al. (see Table 1).

Peruvian Children Versus Adults

As in Lucas et al., we found that there was no age difference in the disjunctive condition, as children and adults were just as likely to label D a flipo, $p = .122$, Fisher's exact test. In the conjunctive condition, in contrast, Peruvian children were more likely to call the ambiguous object D a flipo than Peruvian adults, $p < .001$, Fisher's exact test. Similarly, in the baseline condition, children were more likely than adults to label object D as a flipo, $p < .001$, Fisher's exact test. These differences appear to be responsible for the interaction in the model.

Object G

Despite initial demonstrations that flipos are rare, both children ($M = 0.37$, $SD = 0.49$) and adults ($M = 0.40$, $SD = 0.50$) judged the item G to be a flipo at chance (i.e., 0.50) in the baseline condition, $p = .200$ and .212, respectively, by binomial test. This is similar to the pattern in Lucas et al. and suggests that the training trials raised the baseline probability that blocks were flipos.

Discussion

Peruvian children's and adults' judgments were consistent with the developmental pattern observed in the North American sample. Peruvian children appeared to infer the correct specific causal hypotheses from the data pattern in both the conjunctive and disjunctive case, as evidenced by their specific object choices of A, B, and C, and D, E, and F. They also inferred the more abstract disjunctive and conjunctive overhypotheses, as evidenced by the difference between conditions. They correctly differentiated among the three objects across conditions in the training trials and the test trials, which suggests that they were not confused by the data and did not simply have a yes or no bias or answer at chance.

In the baseline condition, they showed an intermediate pattern and did not differentiate among the objects. This may suggest either that they were unsure which overhypothesis applied or that they were confused without the training trials.

Peruvian adults, in contrast, like U.S. adults, tended to report that only object F activated the machine, irrespective of condition. The two age groups performed comparably in the disjunctive condition. But in the conjunctive condition, Peruvian children were more likely than Peruvian adults to endorse the conjunctive principle. Overall, Peruvian children, like the North American sample, provided responses that appeared to be more sensitive to the training data than Peruvian adults and were more likely to endorse a conjunctive principle.

Experiment 2

In Experiment 2, U.S. children from low-income families were compared to U.S. children from middle- to upper-middle-class families. Additional measures were also administered to determine if differences in SES might be related to other moderating factors that might differ between the groups, such as general cognitive abilities, executive functioning, and language comprehension. If we did find SES differences in causal learning tasks, those might reflect something relatively specific to causal learning, such as differences in informal science experience, and so would not be correlated to other abilities. Alternatively, such differences might reflect broader SES differences in children's cognitive abilities, rather than differences particular to causal learning. Executive function and language ability have been shown to be related to the formation of intuitive theories, particularly, "theory of mind" (Astington & Jenkins, 1999; Carlson & Moses, 2001), and so might also be responsible for differences on causal learning tasks and correlate with performance on those tasks. Similarly, if the low-SES children had difficulties with causal learning tasks because of more general cognitive difficulties, then performance on other cognitive tasks such as Piagetian conservation tasks should correlate with performance on the causal learning tasks. On the other hand, children might do well on causal learning tasks in spite of difficulties in general cognitive tasks, executive function tasks, or language tasks, and that pattern would also be informative.

Method

Participants

Two hundred 4- and 5-year-olds from the San Francisco Bay Area participated in Experiment 2. Ninety-six of these children were from low-SES families, of which 49 children participated in the conjunctive condition (M = 4.66, range = 3.95–5.40) and 47 participated in the disjunctive condition (M = 4.67, range = 3.99–5.45). The remaining 104 children were from middle- to upper-middle-class families. Fifty-six of these children were assigned to the conjunctive condition (M = 4.67, range = 4.02–5.74) and 48 children were in the disjunctive condition (M = 4.66, range = 4.01–5.78). An additional 13 children were tested but not included in analysis for failure to complete the task (n = 2), for machine malfunction (n = 4), or for language comprehension issues (n = 7). Given the relatively small number of available Head Start children, the baseline condition was not administered.

Children from low-SES families were recruited and tested at Head Start programs in Berkeley and Oakland, California. Children are only eligible for enrollment in a Head Start program if the family income falls below the federal poverty line, which in 2015 was below $24,250 for a family of four (Poverty Guidelines, 2015). The higher SES sample was recruited from private preschools throughout Berkeley, California. This population differed slightly from the Lucas et al. (2014) sample, which largely consisted of preschools affiliated with the University of California and, as a result, included a substantial number of low-income student parents with subsidized care. In order to ensure strong SES differences, and to replicate the Lucas et al. results with another clearly middle-class sample, recruitment was expanded to include private, unsubsidized preschools. In 2015, the Berkeley population had a median household income of $66,237 (QuickFacts, Berkeley, California, n.d.). The family incomes of the non-Head Start children in the private preschools were probably higher, although official demographic information was not collected.

Materials and Procedure

Children were tested either in a quiet corner of their classroom or in a separate side room. Each session lasted approximately 15 min and included a causal learning task, a number conservation task, and an executive function task.

All children were reported to be fluent English speakers by their teachers, but many of them came from non-English-speaking backgrounds. Upon testing children, we observed varying degrees of English language proficiency. To more systematically assess this, a vocabulary assessment measure was added midway through data collection. Forty-nine Head Start children and 46 non-Head Start children completed an expressive vocabulary task to ensure their comprehension of the task instructions, as well as to measure language development. Participants completed the tasks in the following order: (a) causal learning task, (b) number conservation task, (c) executive function test, and (d) expressive vocabulary test.

Causal learning task. Materials for this task were identical to those described in Experiment 1; however, the procedure varied slightly among two subgroups of children. For 59 Head Start children and 59 non-Head Start children, the task was identical to that described in Experiment 1 (as well as in Experiment 2 of Lucas et al., 2014) with the exception that participants were tested in English and that the baseline condition was omitted. For 37 Head Start and 45 non-Head Start children who participated in the other version, the task was similar but, as in Lucas et al. Experiment 1, used a more streamlined presentation that did not include the initial demonstration that blickets were rare nor included the G item after the training trials. This was motivated by the possibility that the simpler version might involve fewer information-processing demands. However, preliminary analyses comparing the performance between children tested in the two versions of the study showed no significant differences. We therefore collapsed the two experimental versions.

Number conservation task. A Piagetian conservation of numbers task was administered to assess general cognitive skills. Children were shown two rows of varying lengths, each containing five pennies. Responses were scored as correct if the child said that both rows had the same number of pennies after a transformation, and incorrect if he or she said the rows contained different amounts.

Day–night executive function test. To measure children's ability to inhibit responses, we also administered an executive function task similar to the classic Stroop test. This day–night task (Gerstadt, Hong, & Diamond, 1994) consisted of twenty-two 3 × 3 in. cards. Half of the cards depicted a yellow sun, whereas the other half depicted a blue moon.

The experimenter instructed the child to say "day" when shown a card with the moon and say "night" when shown a card with the sun. Next, the experimenter practiced the game with the child using four cards, two of each kind, and provided feedback. Then the experimenter presented 16 test cards in a quasi-random order (8 with the sun and 8 with the moon). Trials were coded as correct or incorrect, and scores were aggregated across the 16 trials.

Kaufman Assessment Battery for Children–II. We administered the Kaufman Assessment Battery for Children, 2nd ed. (KABC–II) to measure expressive vocabulary and act as a proxy for language comprehension (Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005). Children saw drawings of everyday objects and were asked to label these objects in English. Pictures were presented in order of difficulty, and testing was terminated once the child provided four consecutive incorrect answers. Scores reflect the total number of trials administered.

Results

Training Trials: A Versus B Versus C

As in Experiment 1, we first examined children's judgments in the training trials to assess whether they correctly interpreted the causal relations. Participants should infer that objects A and C are blickets, but B is not, in both conditions. In the *disjunctive* condition, *non-Head Start* children were more likely to call objects A ($M = 1.83$, $SD = 0.48$) and C ($M = 1.79$, $SD = 0.50$) blickets than object B ($M = 0.29$, $SD = 0.62$, $ps < .001$ for contrasts between A vs. B and B vs. C using independent t tests). Similarly, *Head Start* children called C a blicket ($M = 1.43$, $SD = 0.74$) more often than B ($M = 0.98$, $SD = 0.79$), $t(92) = 2.815$, $p = .006$, and called object A a blicket ($M = 1.28$, $SD = 0.80$) marginally more often than B, $t(92) = 1.813$, $p = .073$.

In the *conjunctive* condition, both Head Start and non-Head Start participants labeled object A (Head Start: $M = 1.27$, $SD = 0.86$; non-Head Start: $M = 1.32$, $SD = 0.77$) and C (Head Start: $M = 1.43$, $SD = 0.74$; non-Head Start: $M = 1.45$, $SD = 0.69$) as blickets more often than object B (Head Start: $M = 0.84$, $SD = 0.75$; non-Head Start: $M = 0.79$, $SD = 0.78$), A versus B: $t(96) = 2.634$, $p = .010$; $t(110) = 3.670$, $p < .001$, respectively; B versus C: $t(96) = 3.954$, $p < .001$; $t(110) = 4.763$; $p < .001$, respectively; A versus C: $t(96) = -1.009$, $p = .315$; $t(110) = -0.910$, $p = .365$, respectively.

Test Trial: Blicket Judgments

Figure 4 shows the mean proportion of blicket judgments in the test trial separated by condition and by SES.

Test trial judgments: D versus E versus F. A series of McNemar's exact tests examined whether children correctly discriminated between objects in the test trial. If they are reasoning disjunctively, then they should be less likely to say that D is a blicket than F, but should be equally likely to say that D and E are blickets. If they are reasoning conjunctively, they should be equally likely to say that D and F are blickets, but more likely to say that D

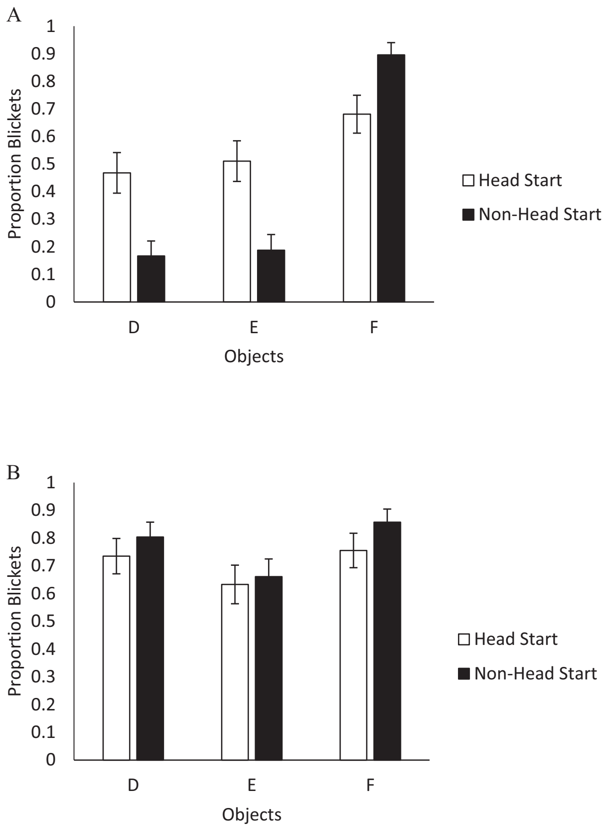

Figure 4. Average proportion of blicket judgments for objects D, E, and F among Head Start and non-Head Start children in the disjunctive condition (A) and conjunctive condition (B). Error bars indicate 1 *SEM* in each direction.

is a blicket than E. If they are confused by the task, then they might fail to differentiate the objects.

In the *disjunctive* condition, *non-Head Start* children called object F a blicket (M = 0.90, SD = 0.31) more often than object D (M = 0.17, SD = 0.38), $p < .001$, and judged D and E (M = 0.19, SD = 0.39), p = 1.0, to be similar. Similarly, *Head Start* children judged object F to be a blicket (M = 0.68, SD = 0.47) reliably more often than object D (M = 0.47, SD = 0.50), p = .044, but did not differentiate between D and E (M = 0.51, SD = 0.51), p = .803.

In the *conjunctive* condition, *non-Head Start* children were equally likely to call objects D (M = 0.80, SD = 0.40) and F (M = 0.86, SD = 0.35) blickets, p = .450. *Non-Head Start* children in the *conjunctive* condition were also significantly more likely to label D a blicket than E (M = 0.66, SD = 0.48), p = .043, indicating that they did not simply respond "yes" to every question.

Head Start children in the *conjunctive* condition were also equally like to label objects D (M = 0.73, SD = 0.45) and F as blickets (M = 0.76, SD = 0.43), p = 1.00, respectively. There were also no significant differences in the Head Start and non-Head Start judgments of all three objects (see below). However, the difference between Head Start children's judgments of objects D and E (M = 0.63, SD = 0.49) did not reach significance, p = .332.

F judgments across conditions. As in Experiment 1, and as predicted, participants generally said that F was a blicket across SES and condition (values ranged from .68 to .90). There was one significant difference: Head Start children were less likely than non-Head Start children to say that F was a blicket in the disjunctive condition, p = .012, Fisher's exact test.

D judgments across conditions. As in Experiment 1, if children had inferred the *conjunctive* principle, then they should say that objects D and F are blickets. If they inferred the disjunctive principle, then they should say that object F is a blicket and D is not.

To test whether SES and condition influenced children's judgments of D, we performed a binary logistic regression with SES, condition, and the interaction between these two as predictors. The resulting model was statistically significant, $\chi^2(3) = 54.433$, $p < .001$, Nagelkerke $R^2 = .319$. Analyses further revealed a main effect of condition, $\chi^2 = 6.911$, $df = 1$, $p = .009$, of SES, $\chi^2 = 9.323$, $df = 1$, $p = .002$, and an interaction between condition and SES, $\chi^2 = 7.729$, $df = 1$, $p = .005$. Below a series of two-tailed Fisher's exact tests further examines responses.

Condition differences. *Non-Head Start* children in the *conjunctive* condition called object D a blicket more often than those in the *disjunctive* condition, $p < .001$. *Head Start* children also judged object D to be a blicket more often in the *conjunctive* than *disjunctive* condition, p = .012.

Head Start versus non-Head Start children. In the *disjunctive* condition, *non-Head Start* children, as noted earlier, were *more* likely to call object F a blicket than *Head Start* children, p = .012, although they were *less* likely to call objects D and E blickets, p = .002 and .001, respectively. This reflected that fact that although both groups discriminated between the two conditions, the *non-Head Start* children did so more dramatically.

In the *conjunctive* condition, in contrast, children enrolled in *Head Start* did not reliably differ from those who were enrolled in *non-Head Start* programs in their judgments of whether objects D, E, and F were blickets, p = .487, .839, and .218, respectively.

Interventions

As an additional measure of children's causal learning, we explored their response to the intervention question, which asked them to choose the object(s) that would activate the machine. We examined first whether children selected multiple objects, the correct response if they had inferred the conjunctive rule, and second whether they selected just object F, the correct response if they had inferred the disjunctive rule (Figure 5 shows the results).

To examine whether children selected multiple objects as a function of SES and condition, we performed a binary logistic regression on their intervention choices. The model, which included SES and condition as fixed factors, was statistically significant, $\chi^2(3) = 37.811$, $p < .001$, Nagelkerke $R^2 = .235$, with a main effect of condition, $\chi^2 = 15.061$, $df = 1$, $p < .001$, and a trending effect of SES, $\chi^2 = 3.456$, $df = 1$, $p = .063$. Overall, children in the conjunctive condition reliably selected multiple objects ($M = 0.55$, $SD = 0.50$) more often than children in the disjunctive condition ($M = 0.20$, $SD = 0.40$), suggesting once again that children correctly inferred the causal structure that was best supported by the evidence.

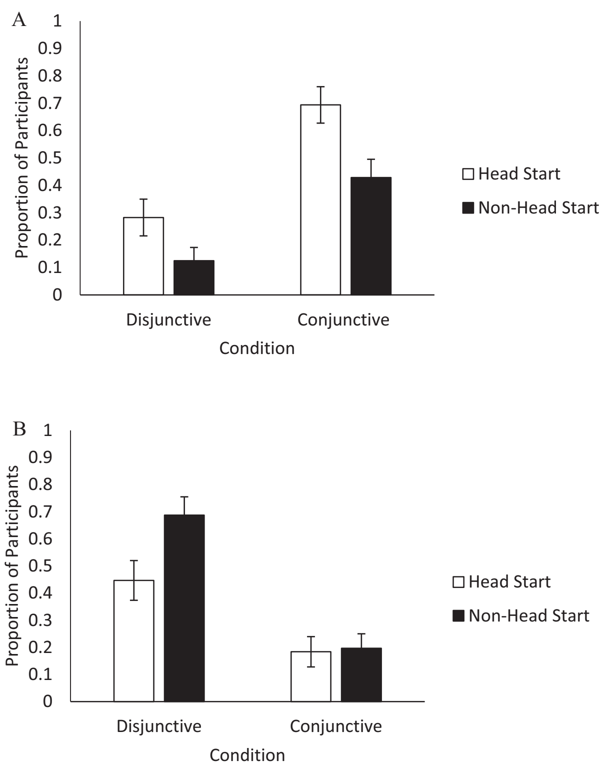

Figure 5. Average proportion of Head Start and non-Head Start children who selected two or more objects (A) or just F (B) when asked to activate the machine. Error bars indicate 1 *SEM* in each direction.

Next, we explored whether SES and condition were a reliable predictor of whether children inferred the correct disjunctive rule (i.e., selecting only object F). To do so, we performed another regression with SES, condition, and the interaction between these two as predictors. This model was significant, $\chi^2(3) = 37.111$, $p < .001$, Nagelkerke $R^2 = .231$, with a main effect of condition, $\chi^2 = 7.352$, $df = 1$, $p = .007$. Children in the disjunctive condition ($M = 0.57$, $SD = 0.50$) were more likely to select just object F, which was the correct disjunctive rule, than children in the conjunctive condition ($M = 0.19$, $SD = 0.40$). However, there was also a main effect of SES, $\chi^2 = 5.485$, $df = 1$, $p = .019$, as non-Head Start children chose only F ($M = 0.42$, $SD = 0.50$) more often than Head Start children ($M = 0.31$, $SD = 0.47$). We found no significant interaction between these predictors, $\chi^2 = 1.954$, $df = 1$, $p = .162$.

As with the blicket judgments, these results suggested that the Head Start children differentiated the two conditions less clearly than the non-Head Start children; they produced fewer just F responses, and more multiple responses in the OR condition. However, when considered separately, the Head Start children, like the non-Head Start children, still made significantly more multiple responses in the conjunctive than disjunctive conditions, $p < .001$, Fisher's exact test, and significantly more just F responses in the disjunctive than conjunctive conditions, $p = .008$, Fisher's exact test, suggesting that they could succeed at the task.

Number Conservation Task

Scores ranged from 0 to 2. An independent samples t test suggests that non-Head Start children ($M = 1.20$, $SD = 0.61$) and Head Start children ($M = 1.00$, $SD = 0.80$) responded differently, $t(175.94) = 1.986$, $p = .049$, on the conservation task.

Day–Night Executive Function Task

According to an independent samples t test, there was a significant difference in performance on the executive function day–night task for Head Start and non-Head Start children, $t(194) = 4.306$, $p < .001$. Non-Head Start children were more accurate at inhibiting their responses ($M = 11.25$ correct of the 16, $SD = 4.10$), than Head Start children ($M = 8.63$ of the 16, $SD = 4.43$).

Kaufman Assessment Battery for Children–II

There was also a significant difference between Head Start and non-Head Start children on their performance on the KABC–II. An independent samples t test revealed that non-Head Start children ($M = 23.07$, $SD = 3.78$) outperformed Head Start children ($M = 16.84$, $SD = 3.94$), $t(93) = 7.855$, $p < .001$.

Relations Between Additional Tasks and Causal Learning Tasks

Despite the differences between SES groups on the conservation, executive function, and vocabulary tests, there were no significant correlations between the executive functioning and vocabulary tasks on any of the crucial blicket judgment or intervention measures (choosing D, choosing multiples in AND, and choosing just F in OR). There was a marginal correlation between conservation and choosing just F in the OR condition ($p = .052$).

Discussion

Despite differences in SES, these findings largely replicate the main pattern described by Lucas et al. (2014). Preschoolers from low-income backgrounds appear to be able to infer both specific causal hypotheses and abstract causal principles. Both groups of children judged the conjunctively active object D to be a blicket more often in the conjunctive condition than in the disjunctive condition. Moreover, all the U.S. children made significantly more multiple object interventions in the conjunctive than in the disjunctive condition, and made more "just F" interventions in the disjunctive condition than in the conjunctive one, paralleling their judgments.

The low-SES and high-SES children performed in a very similar way in the conjunctive condition (see Figure 4B). However, several pieces of evidence suggested that, unexpectedly and unlike the Peruvian children, the low-SES children behaved differently than the higher SES children in the disjunctive condition. The Head Start children were more likely to say that D and E were blickets in the disjunctive condition, and were less likely to say that F was a blicket. They were also more likely to use multiple objects to intervene in the disjunctive condition, and less likely to select just F.

General Discussion

Overall, findings suggest stronger commonalities than differences in causal learning across culture and class contexts. Children from all the groups, 290 children in all, demonstrated some ability to learn both specific causal hypotheses and general causal frameworks as evidenced by their discrimination among the different objects and between the two conditions. In contrast, adults in Peru, like the U.S. adults, did not significantly differentiate between the three conditions. These results appear to be robust and replicable.

One alternative potential explanation for the difference across conditions, in particular, should be considered, however. It was possible that the children did appropriately make disjunctive inferences in the disjunctive case, either because of the training or because that was their default assumption. This in itself would imply some significant cross-cultural competence in causal learning. However, in the conjunctive case they might have simply became confused and showed a "yes" bias, saying that all the objects were "blickets" rather than inferring the conjunctive structure correctly.

Several pieces of evidence weigh against this alternative hypothesis. First, consider the middle-class U.S. children. Replicating the results of Lucas et al., they correctly differentiated between A, B, and C in the conjunctive training trials, and between D and E in the conjunctive test trials. Like the Lucas et al. children they also correctly intervened, producing multiple object interventions in the conjunctive condition, versus just F interventions in the disjunctive one.

The Peruvian children also correctly differentiated between the three choices in the conjunctive, training, and test trials, in fact, as noted their judgments were indistinguishable from those in Lucas et al. However, due to the linguistic differences it was not possible to assess their interventions, and so the evidence was less strong for them.

The Head Start children, in contrast, did show the correct pattern in their interventions, producing multiple object interventions in the conjunctive condition and just F interventions in the disjunctive one. They also significantly differentiated A, B, and C in the conjunctive training trials, though less clearly than the non-Head Start children. However, they did not significantly differentiate D and E in the conjunctive test condition, although their blicket judgments were also not significantly different from those of the middle-class children. So again, the evidence was less strong for them. Further studies would be necessary to definitively rule out this possibility for the Peruvian and Head Start children, but it appears to be less likely than the alternative possibility: that children are correctly inferring the conjunctive structure.

In terms of the predictions, there was some evidence for the Head Start children that difficulties with inhibition and information-processing might have influenced their performance. However, this involved SES differences on the disjunctive task rather than the conjunctive one. Although they appeared to infer the structure correctly, they did so less clearly than the higher SES children, and this might reflect the inhibitory demands of the task. There was no evidence for such differences between the Peruvian and American children. There was also little evidence that the children defaulted to either the disjunctive or conjunctive structure—all the groups of children differentiated between the two training conditions.

Similarly, there was no direct evidence for cross-cultural differences in analytic versus holistic styles of reasoning. However, there was a cultural difference in children's E judgments in the baseline condition, and both Peruvian children and adults made more multiple interventions than U.S. participants. This raises the possibility that Peruvians could be more open to a conjunctive principle than U.S. participants. However, the linguistic differences made the interventions very difficult to interpret and with lack of support from other measurements, it is hard to tell if this finding is actually indicative of such a difference. Additional studies should explore this further. Overall, however, the patterns were strikingly similar in children and adults from a wide range of cultural and economic backgrounds, and suggest that such causal learning abilities may be a feature of young children in general.

The first, most striking, and most practically important result of these studies is that low-income children in Peru and in Head Start programs are able to perform causal learning tasks. These children were certainly able to infer the correct structure in the disjunctive case, and also appeared able to infer the conjunctive structure and to differentiate between the two conditions, suggesting that they could infer "overhypotheses" as well as specific hypotheses. In the case of the low-SES U.S. children, this was true in spite of lower conservation, KABC–II and executive function scores. This, in turn, suggests that these children have basic inferential capacities that could be leveraged in science education.

However, the results also point to some of the difficulties and complexities of such cross-cultural and cross-SES comparisons. The linguistic differences between Spanish and English made it difficult to design comparable intervention instructions in Experiment 1. In Experiment 2, differences in information-processing may have influenced the Head Start children's responses in the disjunctive condition. Furthermore, there were some unexpected English language comprehension issues with the lower income U.S. children. Moreover, all these children were growing up in industrial urban centers with enough parental support to ensure that they were enrolled in high-quality preschools. It would be important to see how children with less support would behave. Similarly, it would be important to test children from small-scale agricultural or forager backgrounds. These results do suggest, however, that children's causal learning abilities extend beyond the WEIRD.

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relationship between language and theory-of mind development. *Developmental Psychology*, *35*, 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Avis, J., & Harris, P. L. (1991). Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. *Child Development*, *62*, 460–467. https://doi.org/10.1111/j.1467-8624.1991.tb01544.x

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science*, *16*, 378–384. https://doi.org/10.1111/j.0956-7976.2005.01544.x

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053. https://doi.org/10.1111/1467-8624.00333

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–495. https://doi.org/10.1037/0033-295X.104.2.367

Coley, J. D. (2012). Where the wild things are: Informal experience and ecological reasoning. *Child Development*, *83*, 992–1006. https://doi.org/10.1111/j.1467-8624.2012.01751.x

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual Review of Anthropology*, *40*, 379–398. https://doi.org/10.1146/annurev-anthro-081309-145822

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5–7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153. https://doi.org/10.1016/0010-0277(94)90068-X

Goodman, N. (1955). *Fact, fiction and forecast* (Vol. 74). Cambridge, MA: Harvard University Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31. https://doi.org/10.1037/0033-295X.111.1.3

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, *24*, 87–92. https://doi.org/10.1177/0963721414556653

Gopnik, A., O'Grady, S., Lucas, C., Griffiths, T., Wente, A., Bridgers, S., . . . Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *PNAS*, *114*, 7892–7899. https://doi.org/10.1073/pnas.1700811114

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205–1222. https://doi.org/10.1111/1467-8624.00224

Gopnik, A., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620–629. https://doi.org/10.103/0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108. https://doi.org/10.1037/a0028044

Griffiths, T. L., & Tenenbaum, J. B. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 323–345). New York, NY: Oxford University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29. https://doi.org/10.1038/466029a

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC–II Assessment* (Vol. 94). New York, NY: Wiley.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*, 678–683. https://doi.org/10.1111/j.1467-9280.2005.01 595.x

Laudan, L. (1978). *Progress and its problems: Towards a theory of scientific growth* (Vol. 282). Berkeley and Los Angelas California: University of California Press.

Legare, C. H., & Harris, P. (2016). The ontogeny of cultural learning. *Child Development*, *87*, 633–642. https://doi.org/10.1111/cdev.12542

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284–299. https://doi.org/10.1016/j.cognition.2013.12.010

Medin, D. L., & Atran, S. (2004). The native mind: Biological categorization, reasoning and decision making in development across cultures. *Psychological Review*, *111*, 960–983. https://doi.org/10.1037/0033-295X.111.4.960

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*, 291. https://doi.org/10.1037/0033-295X.108.2.291

Pearl, J. (2009). *Causality*. New York, NY: Cambridge University Press.

Peng, K., & Knowles, E. D. (2003). Culture, education, and the attribution of physical causality. *Personality and Social Psychology Bulletin*, *29*, 1272–1284. https://doi.org/10.1177/014616720325460

Poverty Guidelines. (2015, November 23). *Office of the assistant secretary for planning and evaluation*. Retrieved from https://aspe.hhs.gov/2015-poverty-guidelines.

QuickFacts, Berkeley, California. (n.d.). *United States Census Bureau*. Retrieved from http://www.census.gov/quickfacts/table/IPE120215/0606000

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, *109*, 211–223. https://doi.org/10.1016/j.cognition.2008.07.017

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, *10*, 322–332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*, 427–442. https://doi.org/10.1111/j.1467-8624.2006.00880.x

Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development*, *84*, 443–454. https://doi.org/10.1111/j.1467-8624.2012.01865.x

Sim, Z., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from two- and three-year-old children. *Developmental Psychology*, *53*, 642–651. https://doi.org/10.1037/dev0000278

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, *10*, 298–306. https://doi.org/10.1111/j.1467-7687.2007.00589.x

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction and search, Springer lecture notes in statistics* (2nd ed.). Cambridge, Massachusetts: MIT Press. (Original work published 1993)

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. https://doi.org/10.1126/science.1192788

Waismeyer, A. S., Meltzoff, A. N., & Gopnik, A. (2014). Causal learning from probabilistic events by human infants: An action measure. *Developmental Science*, *18*, 175–182. https://doi.org/10.1111/desc.12208

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*, 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Xu, F., & Kushnir, T. (Eds.). (2012). *Rational constructivism in cognitive development. Advances in child development and behavior,* (Vol. 43). Waltham, MA: Academic Press.