PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Wente A, Gopnik A, Fernández Flecha M, Garcia T, Buchsbaum D. 2022 Causal learning, counterfactual reasoning and pretend play: a cross-cultural comparison of Peruvian, mixed- and low-socioeconomic status U.S. children. *Phil. Trans. R. Soc. B* 20210345.
https://doi.org/10.1098/rstb.2021.0345

Received: 17 December 2021
Accepted: 28 July 2022

One contribution of 17 to a theme issue 'Thinking about possibilities: mechanisms, ontogeny, functions and phylogeny'.

Subject Areas:
cognition, behaviour

Keywords:
cross-cultural, counterfactual reasoning, causal reasoning, pretend play

Author for correspondence:
Daphna Buchsbaum
e-mail: daphna@brown.edu

Electronic supplementary material is available online at rs.figshare.com.

THE ROYAL SOCIETY PUBLISHING

Causal learning, counterfactual reasoning and pretend play: a cross-cultural comparison of Peruvian, mixed- and low-socioeconomic status U.S. children

Adrienne Wente[1], Alison Gopnik[1], Mari Fernández Flecha[2], Teresa Garcia[1] and Daphna Buchsbaum[3]

[1]Department of Psychology, University of California at Berkeley, Berkeley, CA 94720, USA
[2]Departamento de Humanidades, Pontificia Universidad Católica del Perú, Lima, Peru
[3]Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI 02912, USA

AG, 0000-0002-7183-0867; DB, 0000-0002-8716-7756

Pretend play universally emerges during early childhood and may support the development of causal inference and counterfactual reasoning. However, the amount of time spent pretending, the value that adults place on pretense and the scaffolding adults provide varies by both culture and socioeconomic status (SES). In middle class U.S. preschoolers, accuracy on a pretense-based causal reasoning task predicted performance on a similar causal counterfactual task. We explored the relationship between cultural environment, pretense and counterfactual reasoning in low-income Peruvian ($N = 62$) and low-income U.S. ($N = 57$) 3- to 4-year olds, and contrast findings against previous findings in an age matched, mixed-SES U.S. sample ($N = 60$). Children learned a novel causal relationship, then answered comparable counterfactual and pretense-based questions about the relationship. Children's responses for counterfactual and pretense measures differed across populations, with Peruvian and lower income U.S. children providing fewer causally consistent responses when compared to middle class U.S. children. Nevertheless, correlations between the two measures emerged in all populations. Across cohorts, children also provided more causally consistent answers during pretense than counterfactually. Findings strengthen the hypothesis that causal pretend play is related to causal counterfactual reasoning across cultural contexts, while also suggesting a role for systematic environmental differences.

This article is part of the theme issue 'Thinking about possibilities: mechanisms, ontogeny, functions and phylogeny'.

1. Causal learning, counterfactual reasoning, and pretend play: a cross-cultural comparison of Peruvian, mixed- and low-socioeconomic status U.S. children

Children from a variety of backgrounds spontaneously engage in pretend play[1]. Pretend play universally emerges in typically developing human children between 18 and 24 months of age, and the frequency and complexity of pretense increases during early childhood (for reviews see [3,4]). Even children from cultures where adults discourage pretense spontaneously pretend [3,5], suggesting that pretense may have an innate and evolutionary basis. This may seem perplexing, given that there are no obvious adaptive benefits of pretending. In fact, superficially, it would seem a waste of time to engage in a fictional world when children still have so much to learn about the real one.

However, some researchers have argued that pretend play scaffolds other cognitive skills (e.g. [6]) including theory of mind, divergent thinking and counterfactual reasoning[2]. If these hypotheses hold, pretense would not only be

universal and innate, but arguably provide adaptive benefits. Specifically, one role for pretense might be in facilitating counterfactual causal reasoning. By contrast, other researchers have argued that extensive pretend play may not play a significant, or culturally universal, role in learning [9]. In this paper, we leverage cross-cultural and cross-socioeconomic status (SES) data to test the universality of a relationship between causal counterfactual reasoning and pretense in preschool-aged children.

While children spontaneously engage in pretense by age 2, there are mixed findings as to when children can reason counterfactually, with some procedures showing competence by age 4, and others not until 6 or 8 years of age or even much older (for a recent review see [10]). While some of this variance can likely be explained by researchers leveraging different definitions of counterfactual reasoning (see footnote 2), the complexity of specific experimental procedures or the use of different types of counterfactuals (see [11,12] for further discussion), in all cases counterfactual competence appears to emerge later than pretense abilities.

A number of experiments have reliably shown that 4-year olds can accurately reason counterfactually about deterministic physical causal systems ([11,13,14]; see [15] for related results with older children). For example, both Buchsbaum *et al*. [13] and Nyhout *et al*. [11] used 'blicket machine' style tasks (see [16]) where children were introduced to a novel machine and learned to activate it by placing a specific block on the machine. Then children answered counterfactual questions about this newly learned physical causal relationship. Nyhout *et al*. [11] found that, in a 'blicket machine' task, preschoolers were able to correctly reason about overdetermined scenarios, where an outcome has multiple causes (e.g. would the machine still have activated if only one of two 'blickets' was removed), despite struggling with overdetermined counterfactual scenarios until much later ages in other contexts (e.g. [11,12,17]).

Most relevant to the present paper, Buchsbaum *et al*. [13] argue that pretense and counterfactual reasoning exercise the same underlying causal reasoning mechanisms (see [18,19] for similar arguments) and provide correlational evidence to support this hypothesis. As noted above, the content of counterfactual questions appears to play an important role in children's counterfactual competence. The content of pretend play can also vary widely, reflecting children's differing domain knowledge. In order to help control for potential differences in content and domain knowledge, Buchsbaum *et al*. compared children's responses to a specific counterfactual and pretense prompt in exactly the same situation—a blicket machine-like 'birthday machine'—and so could ask the children to produce pretense responses that were highly comparable to the counterfactual ones.

In two experiments, children learned a novel causal relationship: that a specific block, the zando, activated a zando machine and caused it to play 'Happy Birthday'; however the other block, the non-zando, did not activate the machine. Then children were asked counterfactual questions about the causal relationship between the blocks and the machine (e.g. 'What would happen if the zando was not a zando? Would the machine play music or would it not play music?'). Even 3- and 4-year olds can sometimes correctly answer questions about such machines. After this, the experimenter engaged children in a pretense sequence. During the pretense phase, children were asked comparable questions about a pretend zando machine, a pretend zando and a pretend non-zando.

Buchsbaum and colleagues found that children were more successful at answering hypothetical questions framed within causal pretense, than matching questions framed counterfactually, and argued that children may be more successful at reasoning about alternatives to reality in a pretense context. They also found a correlation between the inferences children made while engaged in pretense and those that they made while reasoning counterfactually. Children who, for example, stated that the pretend zando would activate the pretend machine, also provided correct answers to the counterfactual reasoning questions. This was true even after controlling for age, executive functioning, which is thought to play a role in counterfactual reasoning (e.g. [8,20]), and conservation of number, a cognitive task on which children improve during this same time period, but which is not hypothesized to be related to counterfactual reasoning, to help control for developmental differences in cognitive abilities not specific to counterfactual reasoning or pretense. Together, this supports the hypothesis that causal reasoning during pretense may draw upon the same emerging cognitive capacities as counterfactual causal reasoning, and perhaps that early emerging pretend abilities may even support the development of later counterfactual reasoning abilities.

Intuitively, such a link might exist because both pretend play and counterfactual reasoning are quite similar—both require the ability to 'quarantine' reality in order to reason about alternative fictional possibilities and their outcomes [2,3,21–23]. For example, while pretending, a child may imagine that a banana is a telephone, then work through various instances of what would happen if it were indeed a telephone rather than a banana. This is structurally similar to thinking through counterfactual scenarios, where a hypothetical change to a premise may bring about a range of different outcomes, and an agent must reason about the various possibilities.

As discussed above, researchers have found that children spontaneously pretend at a much younger age, around 18 months, than they first demonstrate competence in counterfactual reasoning, around age 4. In fact, the experimental literature suggests that placing hypothetical questions within a pretense or fantasy setting actually enhances young children's ability to reason from a false premise [24–27]. In these studies, pretense may have supported children's ability to quarantine a fictitious premise from real-world knowledge and reason through causal outcomes. Taken together this could suggest that the ability to reason causally during pretense emerges early and supports further development of complex real-world counterfactual reasoning.

While pretense is ubiquitous in young children regardless of culture and economic standing, empirical work suggests that the sociocultural context substantially shapes children's day to day pretense activities. Several cross-cultural observational studies have documented variance in pretend play along several dimensions. These dimensions include the amount of time that children spend engaged in pretense, the value that parents and caretakers place on pretend play, the extent that parents engage in pretense along with their children, the types of partners young children pretend with (e.g. same-aged children, adults, older children and alone), the subject matter of pretense and the types of props that are used while pretending [5,28–39].

In one study, Callaghan *et al*. [28] sampled Indian, Peruvian and Canadian mother–child dyads, and measured

maternal beliefs about pretense, observed mother–child interactions, and children's tendencies to pretend, both spontaneously and when interacting with an adult who was pretending. In this study, all Canadian mothers reported pretending with their children; however, only 42% of Peruvian mothers and 24% of Indian mothers reported doing so. Callaghan *et al*. next measured both the spontaneous pretense actions of children across cultures and children's response to an experimenter's pretense action. North American children performed more spontaneous pretense acts than either the Indian and Peruvian children, and were more likely follow the experimenter's pretense action with a subsequent pretense action. Furthermore, the average age when children did so was about 1-year younger in the North American sample (34.4 months), than in either the Indian (46.5 months) or Peruvian samples (45.8 months). In the present paper, we leverage this work by comparing Peruvian children to children from the U.S. on pretense and counterfactual causal reasoning tasks.

The early cognitive developmental literature (though see [40] for a critical review of early findings), as well as more recent findings, suggest that higher SES children spend more time pretending than low-SES children, although children from all economic brackets pretend. For example, Doyle *et al*. [41] found that middle class 5- to 7-year-old North American children spent more time in social pretense and pretended for longer periods of time when compared with low-SES North American children. There are similar findings for children in Brazil [36], and in Israel and South Africa [42].

This naturally occurring variance raises questions about how socioeconomic and cultural differences in pretend play may relate to other aspects of cognitive development, such as counterfactual reasoning specifically. In the present paper, we explore this by extending the Buchsbaum *et al*. [13] paradigm to lower income 3- and 4-year-old children in Peru as well as lower income 3- and 4-year-old U.S. children enrolled in Head Start programs. These samples are directly compared to the middle class U.S. sample from Buchsbaum *et al*. [13], Experiment 2.

In addition to the findings discussed above, our own pre-existing observations of the specific preschools we partnered with in Peru during this research suggest that teachers and adults may provide less scaffolding for pretense than do adults in the U.S. For example, the traditional preschool environment in Peru does not provide children with props or space to pretend in, and this is commonly provided to children in the USA, including in the lower income Head Start preschools we also partnered with. Researchers also previously noted that in the Peruvian classrooms, more time is spent in teacher led group activities, and less in self-directed or choice-based free play. This could indicate that children have fewer opportunities to enter into pretense during the school day, and that (consistent with the [28] findings) pretense is less supported by adults. Previous researchers have found that classroom environment impacts children's tendency to engage in pretense [43].

This raises questions about the extent to which the Buchsbaum *et al*. findings will replicate in the Peruvian and low-SES U.S. samples. First, will the correlation between pretense and counterfactual reasoning, and comparatively superior performance on pretense versus counterfactual reasoning tasks, replicate in a context where pretense activities may differ? If causal reasoning during pretense and in counterfactual scenarios is underpinned by the same cognitive capacities, then the correlation should replicate regardless of cultural or socioeconomic background. Moreover, if reasoning about imagined causal scenarios is easier in pretense, perhaps facilitating later counterfactual reasoning, then we might see improved performance on questions couched in a pretense scenario versus equivalent counterfactual questions. On the other hand, if causal counterfactual reasoning and similar reasoning during pretense are not underpinned by the same abilities, but instead reflect some other common factors that are particular to the middle class U.S. children in the Buchsbaum *et al*., sample, then we could see a decoupling of this correlation in the Peruvian and/or low-SES samples.

Additionally, there could be differences in developmental timelines across the three cohorts in both or either skill set. We do not know when low-SES U.S. children and Peruvian children can reliably reason counterfactually, or reason causally while pretending. Given the previous research, it is possible that the Peruvian and low-SES children engage in less pretense than the middle class North American children, and that these differences could be reflected in difference in performance on experimental measures. In particular, we might predict that the differences in everyday pretense activity would lead to differences in reasoning about pretense in a more controlled experimental setting. If this type of pretense reasoning is also related to counterfactual reasoning, we might see differences in both these types of reasoning and might expect that the two types of abilities would be correlated. If these two skill sets are not underpinned by the same cognitive capacities, then we may only see cultural variation in the timeline for the pretense-based task, as that is the task that would presumably be more related to day to day pretense activities, but not necessarily during the counterfactual task. Alternatively, cultural and SES differences in pretend practices may have no effect on the development of causal reasoning either in pretense or when reasoning counterfactually. If so, we might not see any differences across our samples.

2. Methods

The overall structure, stimuli and methods of the experiment replicate those used in Buchsbaum *et al*. [13]. As in Buchsbaum *et al*., we used the stroop-like day/night task [44], to control for individual differences in inhibitory control. As discussed above, the inhibitory control task was included because previous research has found a direct relationship between executive functioning and counterfactual reasoning. As in Buchsbaum *et al*., we also included the Piagetian conservation of number task, as a measure of general cognitive development, similar to numerical age but skill based rather than age based. This was meant to control for non-specific improvement across cognitive tasks and general cognitive development, and for factors like the willingness and motivation to engage in experimental tasks, which should affect these tasks as well as our target tasks. These two measures combined allow us to partial out potential covariance unrelated to the direct relationship between causal reasoning during pretense and causal counterfactual reasoning.

(a) Participants

Participants included 62 3- and 4-year-old Peruvian children, and 57 low-SES U.S. children. This sample size was similar to that used by Buchsbaum *et al*.'s [13] Experiment 2, $N = 60$ 3- and 4-

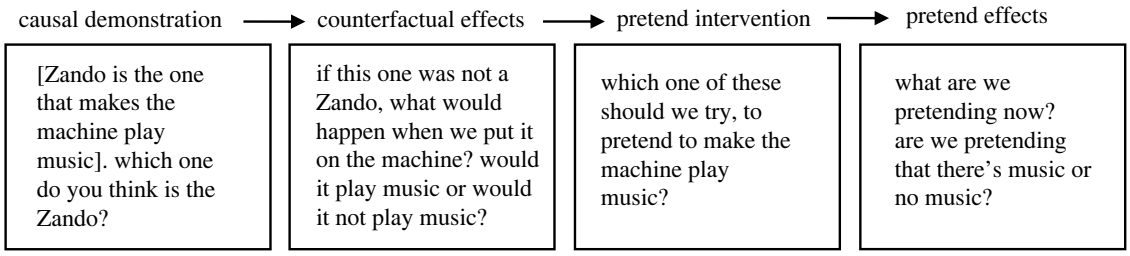

Figure 1. Overview of the Monkey's birthday tasks (full script in the electronic supplementary material, table S1).

year olds. In Peru, the mean age was 4.34 (range = 3.38 to 4.99); for the low-SES U.S. children the mean age was 4.25 (range = 3.26 to 4.97). In Buchsbaum *et al*.'s Experiment 2, children's mean age was 3.89 (range = 2.71–4.95). Thus, the Peruvian and low-SES samples averaged slightly (but significantly) older than the Buchsbaum *et al*.'s Experiment 2 sample (details given in the results section). This slight age difference could have caused the Peruvian and low-SES U.S. children to score higher on experimental measures, but this would lead to the opposite pattern than the hypotheses outlined in the introduction of this paper. An additional six children in the Buchsbaum *et al*., sample, and four children in the low-SES sample failed to learn the causal relationship in the causal demonstration phase of the pretense task even after a second demonstration and were excluded from analysis.

(b) Demographics of Peruvian children

Peruvian children were recruited and tested at Innova schools in Lima, Peru. Innova schools is a chain of private schools that caters to families from the emerging middle class. Schools are frequently located in lower income neighbourhoods in central Lima, or in the outskirts of the city. Most families are internal immigrants from the Andean highlands and come from indigenous backgrounds. Many families own small businesses, and few parents are college educated. Families, however, are financially upwardly mobile and heavily invested in their children's education; families at Innova schools spend around 25% of their income on their children's education. See electronic supplementary material for additional demographic information for the Peruvian cohort.

(c) Demographics of low-SES U.S. children

Low-income U.S. children were recruited and tested in Head Start schools in the San Francisco Bay area. To enroll in Head Start, children must be from families that are considered low income according to the federal poverty guidelines, which at time of retrieval was stated as $26 400 for a family of four (https://www.benefits.gov/benefit/1899 retrieved 3 November 2021).

(d) Demographics of mixed-SES U.S. children

Buchsbaum *et al*. [13] sampled children primarily from U.S. middle and upper middle class backgrounds that reside in the San Francisco Bay Area. Children were tested in university affiliated preschools, science museums or other private preschools. According to 2019 Bay Area census data, the median annual income was $99 406 in.Alameda County and $112 449 in.San Francisco County. (https://www.census.gov/quickfacts/fact/table/alamedacountycalifornia,sanfranciscocountycalifornia/EDU685219 retrieved 3 November 2021).

There was some degree of SES variation across the Peruvian and U.S. samples in addition to cultural variation, with the Peruvian children being from a low-SES background than the U.S. mixed-SES children, although it is difficult to quantify this across countries, especially when considering families 'relative' economic standing within their own community. When contrasting the Buchsbaum *et al*. sample to the Head Start sample, SES is a key differentiating variable.

3. Stimuli and procedure

Participants were tested in a private area in their school. Participants completed the 'Monkey's birthday' task, the day/night stroop task and the Piagetian conservation of number task. Task order was counterbalanced as in Buchsbaum *et al*. [13]. See electronic supplementary material for additional stimuli and counterbalancing details, and for information regarding translation of the experimental prompts from English to Spanish, and procedures to ensure comparability of experimental procedures across samples.

4. Monkey's birthday

(a) Introduction

Figure 1, participants were introduced to a stuffed monkey, named Monkey, or Mono in Spanish. They were told that it was Monkey's birthday. Then the experimenter placed the monkey under the table so that the participant and the experimenter could prepare a surprise for Monkey's birthday. Next the experimenter introduced children to the 'zando' machine (or 'sando' machine in Spanish), which was similar to a blicket machine (see [16]), and was constructed from a wooden box. A doorbell was placed inside, and it played the 'Happy Birthday' song when the experimenter activated it via a button hidden under their foot. The experimenter told children that the zando machine played the song 'Happy Birthday'. They suggested that they use the machine to sing 'Happy Birthday' to Monkey.

(b) Causal demonstration

Next, the experimenter introduced children to the 'zando' and the 'non-zando' blocks. The experimenter told children that the machine only plays 'Happy Birthday' when the zando block is placed on top of the machine, and that they need to figure out which of the two blocks is the zando. To do so, the experimenter placed each block on the machine individually two times. The machine only activated when the zando block was placed on it. Following this initial demonstration, the experimenter asked children which of the two blocks was the zando. If children answered incorrectly, the experimenter repeated the initial demonstration, then again asked which block was the zando. Children had to correctly identify the zando on this second trial before continuing the experiment. After children correctly identified the zando, the experimenter prompted participants to place each of the

blocks on the machine individually. The machine again only activated when the zando block was placed on top. The experimenter and child practiced singing 'Happy Birthday' when the machine activated.

(c) Counterfactual causal effects

Next, the experimenter put the blocks back down on either side of the machine and asked the child to answer two counterfactual causal effects questions about the newly learned causal relationship between the blocks and the machine. These were, 'If this one (while pointing to the zando) was NOT a zando, what would happen when we put it on the machine?' and 'If this one (while pointing to the non-zando) WAS a zando, what would happen when we put it on the machine?'[3]. If children failed to provide an answer, or gave an ambiguous or irrelevant answer, the experimenter asked a forced choice question, which was, 'Would the machine play music or would it not play music?'. Children were scored on whether they answered these questions consistently with the counterfactual identity of the blocks, and counterfactual causal relationships (correct), or with the real-world identity of the blocks and real-world causal relationships (incorrect). After children answered these questions, the experimenter and the child placed each of the blocks on the machine one more time and sang 'Happy Birthday'. Then the experimenter exclaimed that they were ready to sing for Monkey, and they could bring Monkey back.

(d) Transition to pretense

A confederate immediately interjected, and apologetically said that someone else needed to use the machine, and that they needed to bring it to them. Then, the confederate carried the zando machine, and the two blocks, out of the room. The experimenter acted surprised and exclaimed, 'But we didn't sing "Happy Birthday" to Monkey yet! What should we do? Oh, I have an idea! Look what I found! I thought we could PRETEND that this box is my machine! Then we can keep playing!'.

(e) Pretense tasks

The experimenter took out a clear red plastic box (i.e. the pretend zando machine) and placed it in the centre of the table. The experimenter also took out two wooden triangle blocks, one green and one yellow, and placed one on each side of the pretend zando machine. The experimenter said, 'So we can pretend that this box is my machine, and that this block is a zando, and this block is not a zando (while pointing to the blocks). Then we can still sing for Monkey! I'm going to bring Monkey back, ok?'. The experimenter then took Monkey out from under the table and placed him next to the machine.

(i) Pretend intervention questions

Children were next asked two questions about how to pretend that the blocks had the right causal powers. The experimenter prompted children to place one of the two blocks on the machine. To do so, they said, 'So if we're pretending this is my machine, and this is a zando (while pointing to one of the blocks), and this is not a zando (while pointing to the other block), what should we do to pretend to make the machine play music?'. If children failed to provide an answer, or provided an ambiguous or irrelevant answer, they were asked a forced choice question, 'Which one of these should we try, to pretend to make it play music?'. Here, children were asked to generate a causal intervention within the pretend scenario. They could indicate the block that was 'consistent' (the pretend zando) or 'inconsistent' (the pretend non-zando) with the newly learned real-world causal relationship. Children were scored as correct (causally consistent) if they chose the pretend zando and incorrect if they chose the pretend non-zando. Children were not corrected if they chose the pretend non-zando.

(ii) Pretend effects questions

Once children indicated a block, the experimenter placed it on the machine, then asked children a pretend effects question, 'What are we pretending now?'. If children failed to provide a response, or did not give a relevant response, the experimenter asked a forced choice question, 'Are we pretending that there's music or no music?'. In order to directly compare counterfactual and pretend causal reasoning, the pretend effect questions were designed to parallel the earlier questions about the effects of a counterfactual change, where children were asked what would happen if the counterfactual zando and non-zando were placed on the machine. Then, the experimenter suggested they try placing the other block on the machine. They again asked children to indicate what they should pretend was happening, and if they wanted to sing.

Following this, the experimenter suggested they pretend something different. Children were instructed to reverse the causal properties of the blocks (e.g. if the yellow block was originally the pretend zando, then the green block became the pretend zando, and the yellow block became the pretend non-zando). The experimenter repeated the same set of pretend intervention and pretend effects questions stated above. In total, children were asked two pretend effects questions about a pretend zando and two about a pretend non-zando, resulting in four total pretend effects questions. Children were scored on whether they provided a pretend effect response that was causally consistent or inconsistent with the pretense scenario outlined by the experimenter. Children were not corrected if they provided a causally inconsistent response, e.g. stating that placing the pretend zando on the machine did not result in pretend music.

If children respond consistently with the pretend premise, they should state that the pretend zando activates the machine and causes music, while the pretend non-zando does not activate the machine and does not cause music. If, on the other hand, children respond consistently with reality, they should always state that there is no music, given that the effect was pretend—neither block actually causes music to play in the real world. Alternatively, if children's pretense does not correlate with real-world causality, or the pretense scenario set-up by the experimenter, then children should answer randomly. Finally, children may prefer to pretend that both objects are causally effective and always state that there is music, rather than to maintain the real-world causal relationship in their pretense (for instance because they do not transfer the newly learned causal relationship to the pretend scenario, or because they like the song and enjoy pretending that there is music).

(f) Secondary tasks

(i) Open-ended pretense task

Children were asked if they wanted to pretend anything else for Monkey's Birthday. Children were allowed to pretend up to five scenarios. The experimenter and child acted out children's suggestions together.

(ii) Day/night stroop task

The day/night task was modelled after Gerstadt *et al*. [44]. Children were instructed to say 'day' when the experimenter showed them the night card picturing a moon and stars, and 'night' when the experimenter showed them the day card picturing a sun. There was a training phase where children had to provide four correct answers. The test phase consisted of eight of each type of card.

(iii) Conservation of number

Modelled after the classic task developed by Piaget [45], children were shown two rows containing five U.S. pennies a piece. The procedure began with pennies equally spaced across the two rows. The experimenter asked, 'Does this row have more coins? Does this row have more coins? Or do they both have the same number of coins?' while pointing to the appropriate row(s). Then, the rows were expanded and contracted so that one row was longer than the other; however, the two still contained an equal number of pennies. The experimenter gave the same prompt. This was done one more time, resulting in a total of three prompts. Children were scored according to whether they stated that the rows contained an equal number of pennies, or if they believed that one row had more pennies than the other.

5. Results and discussion

The dataset from Buchsbaum *et al*.'s [13] Experiment 2 was obtained, and statistics are presented for all three samples of children. As mentioned in the participants section, mean age differed across the cohorts, $F_{2,\ 176} = 14.05$, $p < 0.001$. Paired contrasts confirmed that the mixed-SES U.S. children averaged slightly younger than both Peruvian, $t_{176} = 5.06$, $p < 0.001$, and low-SES U.S. children, $t_{176} = 3.90$, $p < 0.001$, with no difference between the latter two groups, $t_{176} = 1.07$, $p = 0.54$. Below, we first examine performance on the counterfactual and pretense tasks individually, and then present analyses of the relationship between performance on these two tasks across cohorts.

(a) Causal demonstration

Children successfully learned the novel causal relationship—that the zando activates the machine—with 87% of Peruvian children (54/62), and 91% of low-SES U.S. children (52/57) correctly identifying the zando after the experimenter's first demonstration. This is comparable to the performance of the mixed-SES U.S. sample in Buchsbaum *et al*.'s [13] Experiment 2, with 88% (53/60) of children providing correct responses after the first demonstration, and suggests that children in all three cohorts understood the novel causal relationship. As in Buchsbaum *et al*., we included all children in subsequent analyses. However, excluding children who did not correctly identify the zando after the first demonstration does not change our findings (see electronic supplementary material).

(b) Counterfactual effects performance

Children answered two counterfactual effects questions, resulting in an overall score ranging from 0 to 2. We examined children's counterfactual performance using a linear model with children's age and cohort (Peru, low-SES U.S. and mixed-SES U.S.) as predictors ([13] found that counterfactual performance on this task improved with age in the mixed-SES sample). Counterfactual performance differed by cohort, $F_{2,\ 173} = 12.57$, $p < 0.001$, and performance on the counterfactual task improved with age, $F_{1,\ 173} = 8.87$, $p = 0.003$, with no interaction between cohort and age, $F_{2,\ 173} = 0.63$, $p = 0.53$. Paired contrasts comparing the cohorts confirmed that both children in Peru and low-SES U.S. children performed more poorly than mixed-SES U.S. children and didn't differ from each other, mixed-SES versus low-SES U.S., $t_{173} = 4.28$, $p < 0.001$, mixed-SES U.S. versus Peru, $t_{173} = 4.48$, $p < 0.001$, low-SES U.S. versus Peru, $t_{173} = 0.25$, $p = 0.97$. Thus, despite the overall improvement in counterfactual performance with age across cohorts (figure 2*b*), and the Peruvian and low-SES children's slightly higher mean age, mixed-SES U.S. children provided significantly more consistent responses to counterfactual questions than did the other cohorts, figure 2*a*. In Peru, the mean counterfactual effects score was 0.90 (95% CI = 0.68–1.12), and in the low-SES U.S. cohort, it was 0.91 (95% CI = 0.68–1.15). Children in both of these cohorts were not correct more often than chance, $t_{61} = 0.88$, $p = 0.38$ (Peru; one sample *t*-test), $t_{56} = 0.74$, $p = 0.46$ (low-SES U.S.; one sample *t*-test). By contrast, the mean counterfactual score in the mixed-SES cohort was 1.43 (95% CI = 1.23–1.64). Mixed-SES U.S. children scored significantly above chance on the counterfactual questions, $t_{59} = 4.25$, $p < 0.0001$.

6. Pretense tasks

(a) Pretend intervention

Children from all three cohorts chose to first intervene with the pretend zando (rather than the pretend non-zando) at very high rates (mixed-SES U.S.: 84%; Peru: 84%; low-SES U.S.: 87%), and more often than chance would predict, with no difference between the groups, suggesting that children understood the causal structure of the pretense scenario. More importantly, it demonstrates that children were able to generate a consistent causal intervention within the pretend context (see electronic supplementary material for additional details).

(b) Pretend effects

Children answered four pretend effects questions and were given overall pretend effects scores of 0–4. A linear model with age and cohort as predictors revealed that pretend effects performance differed between cohorts, $F_{1,\ 173} = 7.82$, $p < 0.001$, and performance on the pretense task improved with age, $F_{1,\ 173} = 9.13$, $p = 0.003$, with no interaction between cohort and age, $F_{1,\ 173} = 0.19$, $p = 0.83$. Paired contrasts comparing the cohorts confirmed that both children in Peru and low-SES U.S. children performed more poorly than mixed-SES U.S. children and didn't differ from each other, mixed-SES versus low-SES U.S., $t_{173} = 3.46$, $p = 0.002$, mixed-SES

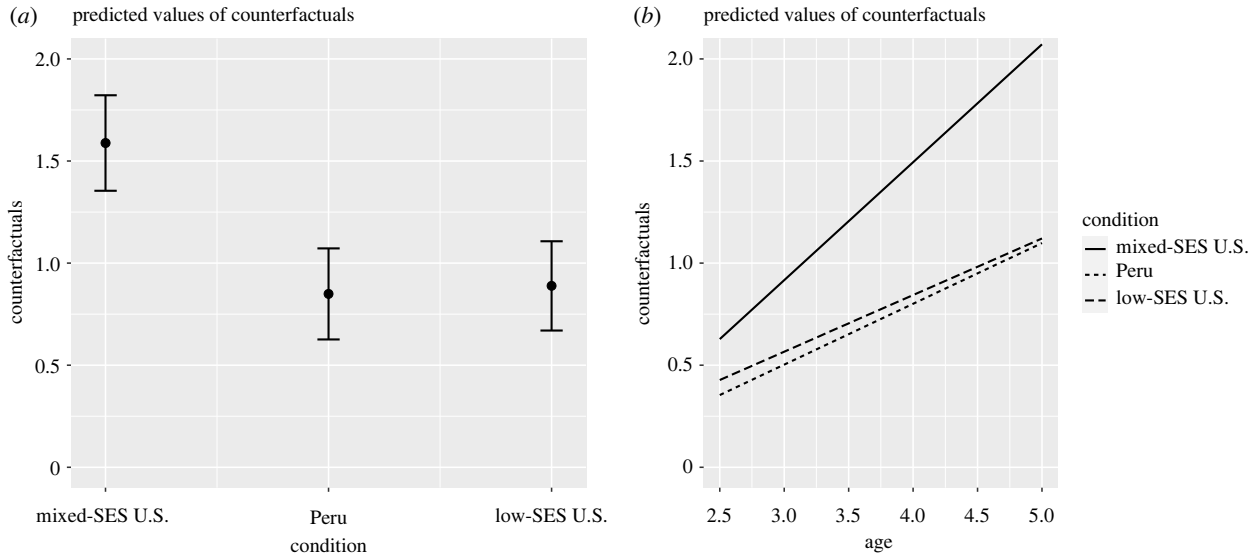

Figure 2. (*a*) Mean counterfactual score (out of 2) and confidence intervals for each cohort. (b) Counterfactual score increases with age across cohorts.

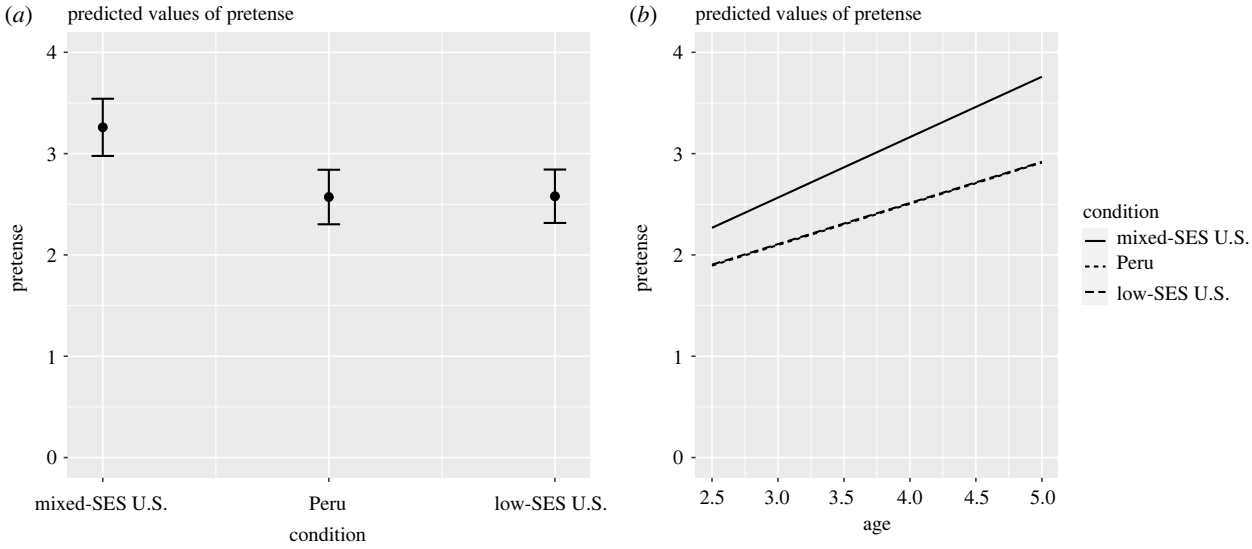

Figure 3. (*a*) Mean pretense score (out of 4) and confidence intervals for each cohort. (*b*) Predicted pretense score increases with age across cohorts.

versus Peru, $t_{173} = 3.46$, $p = 0.002$, low-SES versus Peru, $t_{173} = 0.041$, $p = 0.99$, figure 3*a*. Thus, as in the counterfactual task, despite the overall improvement in pretense performance with age across cohorts (figure 3*b*), and the Peruvian and low-SES U.S. children's slightly higher mean age, they nonetheless performed more poorly than the mixed-SES U.S. cohort collected by Buchsbaum *et al*. [13] on the causal pretense task.

All three cohorts were significantly more likely than chance to give causally consistent pretense answers, in contrast with the counterfactual effects responses of the Peruvian and low-SES U.S. samples (Peru: mean = 2.65; 95% CI = 2.38–2.91; $t_{61} = 4.95$, $p < 0.0001$; low-SES: mean = 2.61; 95% CI = 2.35–2.87; $t_{56} = 4.74$, $p < 0.0001$; mixed-SES U.S. children: mean = 3.10; 95% CI = 2.83–3.37; $t_{59} = 8.09$, $p < 0.0001$). Furthermore, pretense scores did not significantly differ between the initial set of questions and when the roles of the blocks were reversed (see electronic supplementary material). Interestingly, while children in all three cohorts were above chance for both the zando and non-zando blocks, when children did provide an inconsistent answer it was more likely to be a reality-biased one—saying that the zando block does not cause music (which is true in the real-world, but not in the pretend scenario), rather than saying that the non-zando does cause music (see electronic supplementary material).

Taken together, Peruvian and low-SES children's performance on the pretense questions suggests that they are able to reason about a counterfactual premise when it is presented in the context of pretense (whereas they were not above chance when presented with a counterfactual scenario), that they are able to maintain a newly learned a causal relationship within a pretend scenario, and that they flexibly use the pretend roles of the items in order to generate an intervention on the causal system to bring about a pretend outcome.

(c) Comparing counterfactual effects performance and pretend effects performance

Peruvian and low-SES cohorts were successful on the pretend effects task, but at chance on the counterfactual effects task, while the mixed-SES sample was above chance on both tasks. These differences are especially notable given that

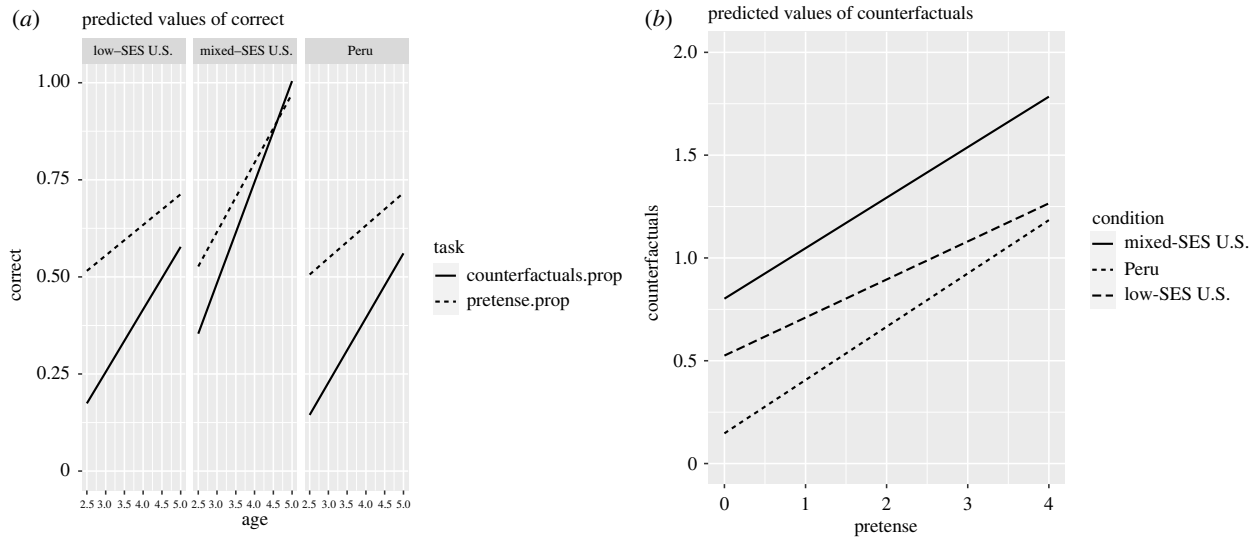

Figure 4. (*a*) Children's relative performance on the counterfactual and pretend effects tasks. (*b*) Higher pretense scores predict higher counterfactual scores, even with age, and performance on the day/night and conservation tasks accounted for.

performance across cohorts improved with age, and the mixed-SES U.S. sample skewed younger than the other samples. We directly compared relative performance on the pretend and counterfactual effects tasks, using a within-subjects mixed-effects linear regression on the proportion of consistent answers, with task and cohort as factors, age as a predictor, and all two-way interactions, and a random intercept for participant. There was a main effect of task, $F_{1,\ 173} = 24.87$, $p < 0.001$. Children answered the pretend questions in a manner consistent with the pretend premise significantly more often than they answered the counterfactual questions consistently with the counterfactual premise. There was also a main effect of age $F_{1,\ 173} = 13.46$, $p < 0.001$, and of cohort, $F_{2,\ 173} = 15.98$, $p < 0.001$, with mixed-SES children outperforming the other two groups on both tasks (as discussed in previous sections), and an interaction between task and cohort $F_{2,\ 175} = 3.21$, $p = 0.043$. There were no other significant interactions, $p > 0.52$, figure 4*a*.

Paired contrasts within each cohort revealed that low-SES U.S., $t_{175} = 3.69$, $p < 0.001$, and Peruvian children, $t_{175} = 4.16$, $p < 0.001$, answered more consistently on the pretend effects questions than on the counterfactual effects questions. This supports the idea that, across cohorts and particularly in the low-SES and Peruvian cohorts, reasoning about an alternative to reality may be easier within pretense than when presented hypothetically. By contrast, while a qualitatively similar trend was present in the mixed-SES U.S. children, it was not significant, $t_{175} = 0.64$, $p = 0.52$, perhaps because their performance on both tasks was already quite high (figure 4*a*). We will return to this point in the general discussion.

(d) Relationship between counterfactual and pretense measures

Finally, we explored whether children who answer the counterfactual effects questions correctly also provide more consistent responses to the pretend effects[4] questions. We used a linear model to explore if pretend effects score predicts children's counterfactual effects score, even when age, day/night score and conservation score are accounted for, and whether this differs across cohorts. We also included interactions between each variable and cohort, to account for age differences between cohorts, and for any differences in performance across the secondary tasks[5]. Both pretend effects score $F_{1,\ 155} = 13.16$, $p < 0.001$, and cohort $F_{2,\ 155} = 6.10$, $p = 0.003$, significantly predicted counterfactual performance (figure 4*b*), with a marginal effect of age, $F_{1,\ 155} = 2.77$, $p = 0.098$. There was no interaction between cohort and pretense performance, $F_{2,\ 155} = 0.12$, $p = 0.88$. None of the other predictors or interactions approached significance, $F \leq 1.76$, $p \geq 0.18$.

In order to examine whether there is a difference in the strength of the relationship between pretend effects score and counterfactual effects score between any of the three cohorts, we compared the slope of the relationship between pretend effects and counterfactual effects scores for all pairs of cohorts. Paired contrasts between cohorts revealed no differences between any pairs of cohorts in the slope of the relationship between pretend effects and counterfactual effects performance, mixed-SES versus low-SES U.S., $t_{155} = 0.38$, $p = 0.70$, mixed-SES versus Peru, $t_{155} = 0.09$ $p = 0.93$, low-SES versus Peru, $t_{155} = 0.47$, $p = 0.64$ (figure 4*b*). Overall, the relationship between the two measures was the same across cohorts.

Finally, we conducted an exploratory analysis to examine the relationship between pretense and counterfactual performance within each cohort. The relationship was significant in the middle class U.S., $t_{173} = 3.29$, $p = 0.001$, and Peruvian $t_{173} = 2.92$, $p = 0.004$ cohorts, with a marginal trend in the same direction for the low-SES sample $t_{173} = 1.71$, $p = 0.089$. This relationship also did not change when only children who identified the zando correctly on the first try during the initial causal scenario were included, middle class U.S., $t_{152} = 2.77$, $p = 0.006$, and Peruvian $t_{152} = 2.86$, $p = 0.005$ cohorts, low-SES sample $t_{152} = 1.89$, $p = 0.060$.

Like the mixed-SES U.S. sample from Buchsbaum *et al.* [13], Peruvian and low-SES U.S. children's ability to consider counterfactual questions about the causal system was predicted by their ability to reason about the causal system during pretense. Pretend effects score significantly predicted counterfactual effects score, and age, general cognitive development as measured by conservation of number, and differences in children's inhibitory control did not account for the relationship. It is worth noting that, when the cohorts were examined individually, the correlation was significant

within the mixed-SES and Peruvian cohorts, but marginal in the low-SES U.S. cohort, and we return to this point in the General Discussion. However, we found no interaction between cohort and pretense score in predicting counterfactual performance, and no differences between any pair of cohorts in the slope of the relationship between pretense and counterfactual performance, suggesting that the pretend effects score is equally predictive of counterfactual effects score across all three cohorts (figure 3). Across cohorts, children who responded consistently to the pretense questions also scored higher on consistency for the counterfactual questions. Taken together, this indicates that the relationship between pretense and counterfactual reasoning is similar across cultures, even though children's average ability to answer these questions differed across cultures.

7. Secondary tasks

(a) Day night stroop task

Children's performance on the day/night inhibition task was significantly predicted by age and did not differ across cohorts, see electronic supplementary material for additional details.

(b) Conservation task

Children's performance on the conservation task was significantly predicted by age, and also by cohort, with the low-SES U.S. cohort outperforming the other two cohorts. See supplementary material for additional details.

(c) Open-ended pretense task

At the end, a majority of children in the Peruvian (77%) and low-SES U.S. (81%) cohorts suggested at least one additional scenario to pretend with the experimenter (this measure was not included in Buchsbaum *et al.*'s [13] Experiment 2, but was included in Experiment 1 with the same population (71%)). There was no difference between the three cohorts, Fisher's Exact Test, $p = 0.38$. This supports the notion that on average children were engaged in the pretend premise and activity. See electronic supplementary material for additional details.

8. General discussion

This study is the first to examine the relationship between causal counterfactual reasoning and pretense outside of a U.S. middle class sample. Across samples, we observed a relationship between the development of causal reasoning during pretense and causal counterfactual reasoning, supporting our primary hypothesis that these two skill sets may be related. Additionally, we saw that, across the age range tested, Peruvian and low-SES children scored lower on measures of counterfactual reasoning and pretense than did middle class U.S. children, supporting the possibility that sociocultural variation in pretense activities may influence the development of abstract causal reasoning more generally. Third, we found higher levels of competence in the pretense-based causal reasoning task than we did in the reality-based counterfactual reasoning task, particularly in the Peruvian and low-SES U.S. cohorts, supporting the possibility that earlier emerging pretense abilities may facilitate later developing counterfactual reasoning abilities. Each of these findings will be further discussed below.

Across samples, children who provided answers during pretense that were consistent with the real-world causal relationship were also more likely to provide accurate responses to the counterfactual questions. The relationship between performance on the pretend effects tasks and counterfactual tasks held even when age, day/night score and conservation score are accounted for. This replicates Buchsbaum *et al.*'s [13] finding with middle class U.S. children and further supports our primary hypothesis that pretense and counterfactual reasoning draw upon the same emerging cognitive capacities.

These findings are particularly remarkable given our second finding, the observed differences across samples: the scores for both the counterfactual and pretense causal reasoning questions did differ across samples, with mixed-income U.S. children scoring more causally consistently than Peruvian and low-SES U.S. children on both counterfactual and pretense questions. Moreover, and strikingly, this did not hold for the secondary executive function and conservation measures. The divergence between the performance on the secondary tasks and the pretense and counterfactual tasks suggests that the cohort differences were not simply the result of differences in general cognitive development, or in engagement or motivation to participate.

We also observed that children, particularly those in the low-SES U.S. and Peruvian cohorts, provided more causally consistent responses to the pretense questions than to the counterfactual questions, suggesting that they may have found it easier to reason about causal scenarios when they were presented as pretense than when they were presented as hypotheticals. In fact, Peruvian and low-SES U.S. children provided causally consistent responses at above chance levels only on the pretense questions, but scored at chance on the counterfactual questions.

Other researchers have found that pretense may enhance young children's ability to reason hypothetically [24–27]. For example, Dias & Harris [24] suggested that the fantasy context may enhance children's ability to 'quarantine' a given false premise from real-world knowledge, which then allows children to make more accurate downstream inferences. Our findings are in line with these claims. By contrast, the mixed-SES U.S. children did not differ significantly in their pretense and counterfactual performance, perhaps because they were already quite successful at both tasks, and close to ceiling at the older end of our age range. As this group may have experienced more opportunity to engage in pretend play early in development, it is possible that these differences in performance are a result of the cultural and socioeconomic environment's impact on pretend play, which we will return to below.

These findings, along with the general developmental timeline in which pretend play and counterfactual reasoning emerge, could imply that the ability to reason hypothetically about alternatives emerges first during pretense, and reality-based hypothetical reasoning emerges later. This raises a compelling question about causation versus correlation: might practicing causal reasoning during pretense actually enhance counterfactual reasoning skills? While our findings are supportive of this possibility, there are reasons to caution against such a strong claim without follow-up research; most importantly, our study design was correlational and not

intended to answer questions about causation. In addition, these results apply to the broader definition of counterfactuals we described in the introduction. Whether pretense correlates with the more specifically defined counterfactuals in other accounts, such as reasoning about nearest possible alternatives to past events, is an interesting open question.

As noted above, the magnitude of the correlation between pretend effects score and counterfactual effects score did not significantly differ across our populations. We found no interaction ($p = 0.88$) and no differences between any of the groups in the slope of the relationship between pretense performance and counterfactual performance ($p \geq 0.64$). However, when the cohorts were examined individually, the relationship between these variables in the low-SES U.S. cohort was marginal. Given the otherwise statistically and qualitatively similar relationship across cohorts, this may be due to the reduction in power when examining the cohorts individually, or due to relatively increased variability within the low-SES U.S. sample. However, we also cannot rule out the possibility of a subtler systematic difference in the strength of the relationship between these variables across cohorts, which our study was not powered to detect. The question of whether not only the existence but the strength of this correlation is consistent across populations is an interesting one and could be pursued in future work with larger samples per cohort.

Further, in order to avoid children responding with respect to a familiar causal schema or script, and to directly compare pretense and counterfactual reasoning skills while controlling for content, our study introduced children to a novel causal system, and then asked them pretense questions about this same system. The similarity between the pretense and counterfactual measures allowed us to directly compare children's performance when asked similar questions about this novel causal system framed hypothetically versus when presented in a pretense scenario. However, this result does not allow us to speak to the generality of this relationship between pretense and counterfactual reasoning, and whether or not pretense writ large may facilitate counterfactual reasoning (a possibility we discuss further below). A longitudinal study, collecting a naturalistic measure of children's pretense activities, and measuring that against counterfactual reasoning performance would be a particularly effective way to address this causal question in future work.

In regard to the difference across samples, the cultural and socioeconomic environment could shape the development of pretense and counterfactual reasoning in a number of ways. For example, children who spend more time pretending may show advanced causal reasoning abilities while pretending and reasoning counterfactually, simply because they have the advantage of increased practice. Furthermore, children from families who value pretend more may spend more time pretending because they are encouraged by adults to do so. However, culture or SES may shape the type of pretense children engage in as well. Engaging in pretense with parents or older children, for example, may scaffold the complexity of children's pretense, and encourage them to think about more complex and varied types of causal relationships. Older children and adults may also correct or shape the inferences that children make during pretense and guide them towards reasoning about relationships that are causally consistent with reality. Of course, it is possible that some other variable accounts for the differences witnessed across samples; this is a question for future research.

The observed cognitive differences across samples, notably, were specific to the counterfactual and pretense causal reasoning tasks. In fact, when contrasting across samples on the available measures, we saw more patterns of similarities than differences. First, we saw that children across samples were equally able to learn the basic real-world cause and effect relationship via observation: that a zando activated the machine, while a non-zando did not. Further, Wente *et al*. [46] conducted a different 'blicket machine' physical causal reasoning experiment with comparable samples of Peruvian, low-income U.S. children and mixed-income U.S. children. In this experiment, all children were equally able to learn much more complex physical causal relationships via observation. This may imply that all children are equally skilled at learning causal structure through observation, and socioeconomic differences are specific to the development of counterfactual causal reasoning skills. This could be further explored through research.

Relatedly, we saw comparable performance across samples on the day/night stroop task and conservation of numbers task. In another study, Zhao *et al*. [47] also found comparability across a similar sample of Peruvian and middle class U.S. children when leveraging an identical conservation of numbers task and executive functioning task. This again, suggests comparability across Peruvian and mixed-SES U.S. samples in these measured facets of broader cognitive development.

The children from Head Start, on the other hand, scored somewhat more accurately than the children from the middle class U.S. and Peruvian backgrounds on the conservation of number task, although comparably on the executive functioning task. One reason for the conservation findings may be the slight difference in ages, with the low-income Head Start children averaging older than the middle class U.S. children; however, replication would be useful before making further claims as these results were unexpected.

Our study design leveraged very similar verbal questions and procedures across pretense and counterfactual reasoning measurements. While this allows us to provide regularities in measurement across skill sets, it may also increase the concern that individual differences in verbal ability could account for at least some of the findings, such as the correlation found across measures. While the present study did control for several aspects of development, including age, conservation of number understanding and executive functioning, it did not explicitly control for verbal abilities (which were also not explicitly measured for the [13] sample).

Although we cannot rule out a possible mediating effect of verbal ability on the relationship between pretense and counterfactual reasoning task performance, previous work has also found a relationship between performance on the day/night task and verbal ability, in children in general and specifically within a HeadStart population [48]. However, in our study, performance on this task did not differ across samples and did not correlate with pretense or counterfactual task performance, suggesting that a broad effect of verbal ability on performance is less likely to explain the differences in performance across the counterfactual and pretense tasks. Nonetheless, it is possible that other common factors, such as verbal or non-verbal IQ, might underpin

the correlation. Measures of verbal ability and IQ would be valuable additional factors to consider in future work.

9. Conclusion

Findings provide support for the hypothesis that causal reasoning during pretend play is related to, and perhaps easier for children than causal counterfactual reasoning. Additionally, findings suggest that the environmental context impacts the development of both types of reasoning. Even so, the relationship between counterfactual reasoning and causal reasoning while pretending was very similar across samples. This suggests that, even if developmental timelines vary by environment, this relationship nonetheless remains stable. Moreover, there has been some debate about benefits of pretend play for other types of cognitive capacities (see e.g. [9]). These findings at least suggest that interventions that encourage pretend play might facilitate counterfactual reasoning, in particular, though this possibility has yet to be tested.

Ethics. The research protocol was approved by the Institutional Review Boards at the University of California Berkeley, protocol numbers 2010-01-631, 2015-03-7243; as well as at Pontificia Universidad Católica del Perú, protocol number N°013-2017/CEI-PUCP. Parents provided written consent prior to their children participating, and children provided verbal assent prior to participating.

Data accessibility. All anonymized and coded data files used for the statistical analyses in this project, as well as the R analysis scripts are available here: https://osf.io/2udb5/.

The data are provided in the electronic supplementary material [49].

Authors' contributions. A.W.: conceptualization, formal analysis, investigation, methodology, writing—original draft and writing—review and editing; A.G.: conceptualization, funding acquisition, resources, supervision and writing—review and editing; M.F.F.: methodology, resources and validation; T.G.: data curation, investigation and project administration; D.B.: conceptualization, formal analysis, methodology, supervision, visualization, writing—original draft and writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This research was supported by funding from the Bezos Family Foundation and Carlos Rodríguez-Pastor.

Acknowledgements. The authors would like to thank Denise Segovia Galván, Emily Anne Shultz and Sophie McMullen for help with data collection and coding, as well as the parents, children, teachers and administrators at Innova Schools in Lima, Peru, and in the YMCA of the East Bay Berkeley and Albany Head Start programs.

EndNotes

[1]We consider pretend play to be play in which the child acts on or in relation to a transformed or alternative reality, intentionally acting 'as if' a non-real state of the world were true [1,2].

[2]Counterfactual reasoning is a wide-ranging interdisciplinary research area. While some researchers use the term 'counterfactual reasoning' to refer to any reasoning about alternative representations that conflict with actual representations of reality, others use the term to refer specifically to reasoning about nearest possible alternatives to past events (see e.g. [7,8]; for some discussion). In this work, we use the broader definition of counterfactual reasoning, meaning any reasoning about alternatives that conflict with reality.

[3]These questions required children to reason about the future effects of present counterfactual identities, in contrast with some previous work which has asked children to reason about the effect of counterfactual past actions on present events.

[4]The counterfactual effects and pretend effects questions directly parallel each other—one asks about the counterfactual causal effect of the block given a counterfactual identity, and one asks about the pretend causal effect of the block given a pretend identity. Therefore, as in [13], we focus our analyses on pretend effects performance. However, see electronic supplementary material for an analysis of the relationship between the pretend intervention performance and counterfactual effects performance.

[5]One child in Peru, four in the low-SES U.S. sample, and two in the Buchsbaum *et al*. [13] sample did not complete one or both of the secondary tasks, and so their data are not included in analyses using these measures.

References

1. Fein 1981.
2. Leslie AM. 1987 Pretense and representation: the origins of 'theory of mind'. *Psychol. Rev.* **94**, 412–422. (doi:10.1037/0033-295X.94.4.412)
3. Lillard A, Pinkham AM, Smith E. 2011 Pretend play and cognitive development. In *The Wiley-Blackwell handbook of childhood development, second edition* (ed. U Goswami), pp. 121–147. Hove, UK: Psychology Press.
4. Nielsen M. 2015 Pretend play and cognitive development. In *International encyclopedia of the social & behavioral sciences*, vol. **18**, pp. 870–875, 2nd edn.
5. Carlson SM, Taylor M, Levin GR. 1998 The influence of culture on pretend play: the case of Mennonite children. *Merrill-Palmer Q.* **44**, 538–565.
6. Weisberg DS. 2015 Pretend play. *Cogn. Sci.* **6**, 249–261.
7. Beck SR. 2016 Why what is counterfactual really matters: a response to Weisberg and Gopnik (2013). *Cogn. Sci.* **40**, 253–256. (doi:10.1111/cogs.12235)
8. Beck SR, Riggs KJ. 2014 Developing thoughts about what might have been. *Child Dev. Perspect.* **8**, 175–179. (doi:10.1111/cdep.12082)
9. Lillard A, Lerner MD, Hopkins EJ, Dore RA, Smith ED, Palmquist CM. 2013 The impact of pretend play on children's development: a review of the evidence. *Psychol. Bull.* **139**, 1. (doi:10.1037/a0029321)
10. Nyhout A, Ganea PA. 2019 The development of the counterfactual imagination. *Child Dev. Perspect.* **13**, 254–259. (doi:10.1111/cdep.12348)
11. Nyhout A, Henke L, Ganea PA. 2019 Children's counterfactual reasoning about causally overdetermined events. *Child Dev.* **90**, 610–622. (doi:10.1111/cdev.12913)
12. Rafetseder E, Schwitalla M, Perner J. 2013 Counterfactual reasoning: from childhood to adulthood. *J. Exp. Child Psychol.* **114**, 389–404. (doi:10.1016/j.jecp.2012.10.010)
13. Buchsbaum D, Bridgers S, Weisberg DS, Gopnik A. 2012 The power of possibility: causal learning, counterfactual reasoning, and pretend play. *Phil. Trans. R. Soc. B* **367**, 2202–2212. (doi:10.1098/rstb.2012.0122)
14. Schulz LE, Gopnik A, Glymour C. 2007 Preschool children learn about causal structure from conditional interventions. *Dev. Sci.* **10**, 322–332. (doi:10.1111/j.1467-7687.2007.00587.x)
15. McCormack, Ho, Gribben, O'Connor, Hoerl. 2018.
16. Gopnik A, Sobel DM. 2000 Detecting blickets: how young children use information about novel causal powers in categorization and induction. *Child Dev.* **71**, 1205–1222. (doi:10.1111/1467-8624.00224)
17. Rafetseder E, Cristi-Vargas R, Perner J. 2010 Counterfactual reasoning: developing a sense of 'nearest possible world'. *Child Dev.* **81**, 376–389. (doi:10.1111/j.1467-8624.2009.01401.x)
18. Walker CM, Gopnik A. 2013 Causality and imagination. In *The Oxford handbook of the development of imagination*, pp. 342–358.
19. Weisberg DS, Gopnik A. 2013 Pretense, counterfactuals, and Bayesian causal models: why what is not real really matters. *Cogn. Sci.* **37**, 1368–1381. (doi:10.1111/cogs.12069)
20. Beck SR, Riggs KJ, Gorniak SL. 2009 Relating developments in children's counterfactual thinking and executive functions. *Thinking Reasoning* **15**, 337–354. (doi:10.1080/13546780903135904)

21. Amsel E, Smalley JD. 2000 Beyond really and truly: children's counterfactual thinking about pretend and possible worlds. In *Children's reasoning and the mind* (eds P Mitchell, KJ Riggs), pp. 121–147. Hove, UK: Psychology Press.
22. Harris PL. 2000 *The work of the imagination*. Oxford, UK: Blackwell Publishing.
23. Hoerl C, McCormack T, Beck SR (eds) 2011 *Understanding counterfactuals, understanding causation: issues in philosophy and psychology*. Oxford, UK: Oxford University Press.
24. Dias MG, Harris PL. 1988 The effect of make-believe play on deductive reasoning. *Br. J. Dev. Psychol.* **6**, 207–221. (doi:10.1111/j.2044-835X.1988.tb01095.x)
25. Dias MG, Harris PL. 1990 The influence of the imagination on reasoning by young children. *Br. J. Dev. Psychol.* **8**, 305–318. (doi:10.1111/j.2044-835X.1990.tb00847.x)
26. Hawkins J, Pea RD, Glick J, Scribner S. 1984 'Merds that laugh don't like mushrooms': evidence for deductive reasoning by preschoolers. *Dev. Psychol.* **20**, 584–594. (doi:10.1037/0012-1649.20.4.584)
27. Scott FJ, Baron-Cohen S, Leslie A. 1999 'If pigs could fly': a test of counterfactual reasoning and pretense in children with autism. *Br. J. Dev. Psychol.* **17**, 349–362. (doi:10.1348/026151099165339)
28. Callaghan T, Moll H, Rakoczy H, Warneken F, Liszkowski U, Behne T, Tomasello M, Collins WA. 2011 Early social cognition in three cultural contexts. *Monogr. Soc. Res. Child Dev.* **76**, 1–142. (doi:10.1111/j.1540-5834.2011.00603.x)
29. Farver JAM, Howes C. 1993 Cultural differences in American and Mexican mother-child pretend play. *Merrill-Palmer Q.* **39**, 344–358.
30. Farver JAM, Kim YK, Lee Y. 1995 Cultural differences in Korean- and Anglo-American preschoolers' social interaction and play behaviors. *Soc. Res. Child Dev.* **66**, 1088–1099. (doi:10.2307/1131800)
31. Farver JAM, Kim YK, Lee-Shin Y. 2000 Within cultural differences examining individual differences in Korean American and European American preschoolers' social pretend play. *J. Cross-Cult. Psychol.* **31**, 583–602. (doi:10.1177/0022022100031005003)
32. Farver JAM, Lee-Shin Y. 2000 Acculturation and Korean-American children's social and play behavior. *Soc. Dev.* **9**, 316–336. (doi:10.1111/1467-9507.00128)
33. Farver JAM, Wimbarti S. 1995 Indonesian children's play with their mothers and older siblings. *Soc. Res. Child Dev.* **66**, 1493–1503. (doi:10.2307/1131659)
34. Gaskins S, Göncü A. 1992 Cultural variation in play: a challenge to Piaget and Vygotsky. *Q. Newsletter Lab. Comp. Cogn.* **14**, 31–35.
35. Göncü A, Mistry J, Mosier C. 2000 Cultural variations in the play of toddlers. *Int. J. Behav. Dev.* **24**, 321–329. (doi:10.1080/01650250050118303)
36. Gosso, Morais, Otta, 2007.
37. Haight WL, Wang X, Fung HH, Williams K, Mintz J. 1999 Universal, developmental, and variable aspects of young children's play: a cross-cultural comparison of pretending at home. *Child Dev.* **70**, 1477–1488. (doi:10.1111/1467-8624.00107)
38. Lancy DF. 2007 Accounting for variability in mother-child play. *Am. Anthropol.* **109**, 273–284. (doi:10.1525/aa.2007.109.2.273)
39. Parmar P, Harkness S, Super CM. 2004 Asian and Euro-American parents' ethnotheories of play and learning: effects on preschool children's home routines and school behavior. *Int. J. Behav. Dev.* **28**, 97–104. (doi:10.1080/01650250344000307)
40. McLoyd V. 1982 Social class differences in sociodramatic play: a critical review. *Dev. Rev.* **2**, 1–30. (doi:10.1016/0273-2297(82)90002-8)
41. Doyle *et al.* 1991.
42. Udwin, Shmukler 1981.
43. Ashiabi 2007.
44. Gerstadt CL, Hong YJ, Diamond A. 1994 The relationship between cognition and action: performance of children 3–7 years old on a stroop-like day-night test. *Cognition* **53**, 129–153. (doi:10.1016/0010-0277(94)90068-X)
45. Piaget J. 1968 Quantification, conservation, and nativism. *Science* **162**, 976–979. (doi:10.1126/science.162.3857.976)
46. Wente AO, Kimura K, Walker CM, Banerjee N, Fernández Flecha MF, MacDonald B, Gopnik A. 2017 Causal learning across culture and socioeconomic status. *Child Dev.* **90**, 859–875. (doi:10.1111/cdev.12943)
47. Zhao X, Wente A, Fernández Flecha MF, Galvan DS, Gopnik A, Kushnir T. 2021 Culture moderates the relationship between self-control ability and free will beliefs in childhood. *Cognition* **210**, 104609. (doi:10.1016/j.cognition.2021.104609)
48. Fuhs MW, Day JD. 2011 Verbal ability and executive functioning development in preschoolers at head start. *Dev. Psychol.* **47**, 404. (doi:10.1037/a0021065)
49. Wente A, Gopnik A, Fernández Flecha M, Garcia T, Buchsbaum D. 2022 Data from: Causal learning, counterfactual reasoning and pretend play: a cross-cultural comparison of Peruvian, mixed- and low-SES U.S. children. *Figshare*.
50. Cook C, Goodman ND, Schulz LE. 2011 Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition* **120**, 341–349. (doi:10.1016/j.cognition.2011.03.003)
51. Harris PL, German T, Mills P. 1996 Children's use of counterfactual thinking in causal reasoning. *Cognition* **61**, 233–259. (doi:10.1016/S0010-0277(96)00715-9)
52. Henrich J, Heine SJ, Norenzayan A. 2010 The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–135. (doi:10.1017/S0140525X0999152X)
53. Schulz LE, Bonawitz EB. 2007 Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Dev. Psychol.* **43**, 1045–1050. (doi:10.1037/0012-1649.43.4.1045)
54. Singer DG, Golinkoff RM, Hirsh-Pasek K. 2006 *Play= learning: how play motivates and enhances children's cognitive and social-emotional growth*. Oxford, UK: Oxford University Press.
55. Schwartzman HB. 1976 The anthropological study of children's play. *Annu. Rev. Anthropol.* **5**, 289–328. (doi:10.1146/annurev.an.05.100176.001445)