CHILD DEVELOPMENT

Child Development, xxxx 2019, Volume 00, Number 0, Pages 1–17

Young Children Are Wishful Thinkers: The Development of Wishful Thinking in 3- to 10-Year-Old Children

Adrienne O. Wente, Mariel K. Goddu, Teresa Garcia, and Elyanah Posner
University of California, Berkeley

María Fernández Flecha
Pontificia Universidad Católica del Perú

Alison Gopnik
University of California, Berkeley

Previously, research on wishful thinking has found that desires bias older children's and adults' predictions during probabilistic reasoning tasks. In this article, we explore wishful thinking in children aged 3- to 10-years-old. Do young children learn to be wishful thinkers? Or do they begin with a wishful thinking bias that is gradually overturned during development? Across five experiments, we compare low- and middle-income United States and Peruvian 3- to 10-year-old children (N = 682). Children were asked to make predictions during games of chance. Across experiments, preschool-aged children from all backgrounds consistently displayed a strong wishful thinking bias. However, the bias declined with age.

Previously, psychologists have found that both adults and young children frequently hold optimistic beliefs. However, the underlying cause of this optimism is less clear. Some researchers have argued, more specifically, for a "wishful thinking bias" also called a "desirability bias." According to this hypothesis, a desire or preference for a specific outcome directly increases the belief that the desirable outcome will occur. While several studies have explicitly measured wishful thinking in adults and older children (see Krizan & Windschitl, 2007, 2009, 2007, 2009 for review), finding some support for this hypothesis, previous studies have not explicitly measured wishful thinking in young children. In this article, we explore wishful thinking in young children, aged 3–10, from Peru and the United States, to uncover the development trajectory of wishful thinking.

[Article updated on August 19, 2019, after first online publication on August 9, 2019: In Table 2, data under "95% CI for the mean" for second youngest U.S participants was corrected from ".29, .34" to ".03, .34."]

The authors thank the Bezos Family Foundation and Carlos Rodríguez-Pastor for funding. They also thank research assistants Denise Segovia Galván, Karen Ramirez Balarezo, and Ingrid Del Rosario Abanto Hurtado for data collection and Sophie McMullen for help with coding.

Correspondence concerning this article should be addressed to Adrienne O. Wente, Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704. Electronic mail may be sent to adrieorchid@gmail.com.

The Relationship Between Preferences and Expectations

For decades researchers have documented a link between preferences and expectations, finding that people often hold expectations that are congruent with their preferences (Granberg & Brent, 1983; Hayes, 1936; Ogburn, 1934). For example, Granberg and Brent (1983) tallied survey data across eight presidential elections and found that four of five U.S. adults believed their preferred presidential candidate would win. While this finding, and others, suggest that people may have optimistic beliefs, it is not always clear why people have these beliefs. Wishful thinking, as opposed to optimism more generally, specifically implies that desires have a causal influence on beliefs. In the example just given, presidential preferences may have driven people's election predictions—they may have believed the candidate would win precisely because they wanted the candidate to win, a classic case of wishful thinking. Alternatively, however, the prediction may have driven the preference; people may have preferred that specific candidate because they believed that candidate would win (not a case of wishful thinking). Finally, a third variable could have driven both their preferences, and their predictions; for example, both predictions and

DOI: 10.1111/cdev.13299

preferences could be shaped by other people's predictions and preferences, also called a "bandwagon" effect' (and not a case of wishful thinking).

The unrealistic belief would only count as wishful thinking if it was directly caused by the desire, so to demonstrate wishful thinking we need to manipulate desires without otherwise changing beliefs. To explicitly test wishful thinking, researchers have used games of chance to experimentally manipulate participants' desires, and then measured the influence of those desires on participants' predictions about stochastic events.

The first "wishful thinking" study was conducted with 9- to 11-year-old children. In this study, Marks (1951) introduced children to a deck of cards, some of which were marked on one side, and told children the percentage of marked cards in the deck. Across conditions, decks contained different percentages of marked cards (10%, 30%, 50%, 70%, and 90%). Across conditions, participants were also told they would win (gain condition) or lose (loss condition) a point if they blindly drew a marked card from the deck. After this, participants were asked to guess which card they thought they would select from the deck. Responses varied according to both the probability and desirability of selecting a marked card. Holding likelihood constant, participants believed they were more likely to select a marked card in the gain conditions than in the loss conditions, suggesting that desirability altered expectations. Children's estimates of drawing their preferred card was heavily skewed across these different ratios. For example, when the probability of drawing a desirable marked card was 5–5, 90% of children believed they would draw the desirable card. When the probability of the desirable outcome was a slim 1–9, 47% of children still believed they would draw the desirable card.

Since this time, several variations of this paradigm have been conducted with adults, but none with younger children. In one of these studies, Irwin (1953) used a nearly identical paradigm to Marks (1951). Irwin found that when the marked card was desirable, 61% of participants (across the various probabilities) stated they would draw a marked card, however, when the marked card was undesirable, only 48% did so, suggesting some effect of desirability on adults' expectations, albeit a much smaller effect than Marks found in children. Meta-analyses drawing upon several similar studies yielded comparable findings (Krizan & Windschitl, 2007, 2009). Contrasting these findings with Marks (1951) implies that a wishful thinking bias may be stronger for school-aged children than it is for adults.

This raises questions about the development of wishful thinking. Perhaps a wishful thinking bias is acquired during early childhood. If this is so, we might predict that the bias would increase with development. Alternatively, desires could initially constrain young children's predictions and children may gradually overcome this bias with age. If so, we should see a desire bias even in very young children, and it should weaken over the course of development.

Developmental Research on Optimism and Positivity

While wishful thinking has not been explicitly measured in young children, several developmental studies have explored optimism more generally. Many of these studies have measured young children's beliefs about trait stability, finding that younger children often exhibit a "positivity bias" when they evaluate trait stability over the course of time—young children expect negative traits to change for the better, but believe that positive traits will remain stable (Diesendruck & Lindenbaum, 2009; Heyman & Giles, 2004; Lockhart, Chang, & Story, 2002; Lockhart, Nakashima, Inagaki, & Keil, 2008). For example, Lockhart et al. (2002) introduced young children (5- to 6-year-olds), older children (7- to 10-year-olds), and adults to a story in which characters wanted to change a negative attribute for the better (e.g., become more athletic, or more attractive). Younger children were likely to believe that these negative attributes would change for the better, whereas adults judged them to be more stable over time. Similarly, other studies have asked children about story characters who wanted to change their positive attribute into a negative one. For example, Heyman and Giles (2004) introduced children to a character who was smart, but did not want to be smart. In these types of scenarios, young children tend to state that the positive trait will persist over time, even when the protagonist wished otherwise.

Similarly, in one study, Boseovski and Lee (2008) introduced children to a story character who either performed positive or negative actions toward another character. In this study, children readily extended positive attributes to the story character after viewing positive actions, however, they were hesitant to make a negative inference after viewing negative actions. Relatedly, Boseovski (2012) explored children's endorsement of an informant's testimony, finding that children were more likely to endorse an informant who stated a person was nice, rather than one who stated they were mean.

While these results suggest that young children often hold positive or optimistic beliefs, it is not clear if young children's *own* preferences caused their responses. For example, while some of these studies explicitly stated a story character's desires (e.g., stated that the story character wanted to change), and experimentally manipulated positive and negative trait valence (e.g., being nice vs. mean), none have measured if the participants' *own desires* actually aligned with the *story character's desires* or the *trait valence* (Heyman & Giles, 2004; Lockhart et al., 2002; Lockhart et al., 2008). As there were no explicit manipulations of participants' desires or measurement of their preferences, it would be a stretch to argue that these findings are evidence for wishful thinking in young children, although they are in line with this hypothesis. Furthermore, there are several alternative explanations for children's optimism in these previous studies. These alternative explanations are discussed in the following section.

In contrast to the studies reviewed earlier, other developmental studies have taken a first-person perspective, finding that young children often exhibit overconfidence in their own abilities across a variety of situations. For example, Parsons and Ruble (1977) found that preschool-aged children expected to do well on a puzzle task, even after being told they were doing poorly; however, 6-year-olds updated their expectations in response to negative feedback. Similarly, Plumert (1995) found that 6-year-olds, but not 8-year-olds, demonstrated overconfidence in their physical abilities, such as running fast. In another study, Lockhart, Goddu, and Keil (2017) found that 5- to 7-year-olds were more likely to think that they would eventually acquire complete knowledge than were 8- to 10-year-olds.

In these first-person studies, it seems quite plausible that children preferred the positive outcome (e.g., running fast or acquiring more knowledge). If so, this preference could have influenced their responses. However, again children's desires were not experimentally manipulated across conditions, nor were they explicitly measured. In addition, as Lockhart et al. (2002) discuss, there are several alternative explanations for these results. These alternative explanations are outlined in the following section.

Alternative Explanations for Young Children's Optimism

Beliefs About the Base Rate Prevalence of Positive Attributes

Experimental evidence suggests developmental differences in children's prior knowledge about personality traits; younger children believe that positive traits are more prevalent than older children (Lockhart et al., 2002). A strong prior belief in favor of positive attributes may cause inferential biases, even after viewing evidence to the contrary (Gopnik, Griffiths, & Lucas, 2015; Seiver, Gopnik, & Goodman, 2013). In this case, it would be important to explore why children have different beliefs about the base rate prevalence of traits. Children may believe that positive traits are common because this belief aligns with their desires (e.g., wishful thinking). However, a number of other factors could shape the development of these beliefs, such as evidence from the testimony of adults.

Beliefs About the Controllability of Traits

Research also suggests that younger children believe people have more control over the development of traits and abilities than do adults (Lockhart et al., 2002; Stipek & Mac Iver, 1989). This may cause younger children to believe that people can improve over time if they want to. Again, wishful thinking could influence children's beliefs about the controllability of traits; young children may believe that people can control outcomes because they wish it to be so. However, young children may believe this for other reasons. In particular, they may encounter first-person or testimony evidence that leads them to conclude this.

First-Person Evidence

Children likely receive different patterns of evidence in their day to day lives than older children and adults, and this could shape their beliefs about controllability and malleability. Indeed, young children's traits and abilities do rapidly change during development, which may lead the children to believe that traits and abilities are quite malleable. Young children may also believe that adults generally have more positive traits than children; for example, adults do run much faster and have acquired a much larger body of knowledge. As a result, children may come to believe that with age everyone's relative standing will improve.

Testimony Evidence

Adults may also selectively provide younger children with positive and encouraging feedback and this may cause young children to develop optimistic beliefs about their own abilities. While kindergarteners generally rate their future academic attainment higher than fourth graders do, Stipek and Daniels (1988) found that kindergarteners who were given salient positive and negative feedback, similar to the feedback fourth graders generally receive, rated themselves comparably to fourth graders. In another study, Stipek, Roberts, and Sanborn (1984) found that 4-year-old children adjusted their estimates of success in response to adult feedback. Both of these studies suggest that testimony evidence does shape children's beliefs about their own abilities. This sort of testimony evidence, rather than wishful thinking, could have underpinned children's confidence in the previous studies.

At least one study provides more direct support for wishful thinking. In this study, Stipek et al. (1984) explored whether 4-year-olds' overconfidence was impacted by incentivizing success. Children were introduced to a challenging task. In an incentivized condition, children were told they would receive a reward for success; children in a control condition were not rewarded for success. After struggling with the task, children's estimates of eventual success remained higher when success was incentivized than when it was not, suggesting that manipulating children's desires (through manipulating the incentive) altered children's expectancies about the outcome. However, again, there is at least one good alternative explanation for this finding; it is possible that through offering an incentive, experimenters also altered children's motivation, which in turn could have rationally influenced their actual likelihood of success and corresponding predictions. In this case, desires would not directly impact children's expectations, but rather their motivation, which in turn could influence their expectations—in other words children might recognize that they were more motivated in the incentivized condition and accurately predict that motivation improves performance. If so, wishful thinking would not be the cause of children's optimism.

There is also research indicating that desires strongly constrain children's initial beliefs about agency. For example, Gopnik and Slaughter (1991) found that preschool-aged children's recollection of their own past desires was often biased by a current desire. In another study, Moore, Jarrold, Russell, Lumb, Sapp, and MacCallum (1995) asked children to infer another person's desire when it was in conflict with their own desire, and experimentally varied the magnitude of participants' desires. They found that only 5-year-olds could accurately predict another person's desire when there was a strong conflict of desire; 3- and 4-year-olds could not. However, when there was not a strong conflict of desire, even 3-year-olds could make accurate judgments. These studies differ from studies on wishful thinking because they ask children to predict mental states, rather than future outcomes that are relevant to participants. However, results could indicate that desires more broadly constrain young children's inferences.

Taken together, research supports the notion that young children frequently hold optimistic beliefs, particularly about traits and abilities. Research also suggests that desires bias young children's ability to accurately predict mental states. However, it is not yet clear if young children engage in wishful thinking, and if desires bias children's predictions about outcomes.

Previous studies have, however, explicitly measured wishful thinking in school-aged children and adults, generally finding a bias when asking participants to make binary predictions about stochastic events. These findings suggest that the bias may attenuate with age. No previous studies have directly tested wishful thinking in young children, and in particular, none have measured if desires influence young children's predictions about stochastic events. In this article, experimenters use games of chance to directly manipulate young children's desires and measure the influence of desirability on probability judgments.

Probability Judgments in Early Childhood

One reason that a Marks (1951) wishful thinking style of paradigm has not been extended to young children sooner is because of the earlier consensus that young children have difficulty understanding probability. In the first of these studies, Piaget and Inhelder (1975) introduced 5- to 12-year-old children to a container holding two colors of chips. The proportion of each color varied. Children were asked to point to the color of chip they believed would be randomly selected. Children under the age of 7 did not provide accurate predictions.

Other studies have challenged this position, showing that under certain conditions young children do demonstrate a basic understanding of probability (Denison & Xu, 2014; Yost, Siegel, & Andrews, 1962). For example, Yost et al. (1962) informed 5-year-old children that they would receive a prize if they randomly selected a specific color of chip from a container. Then children were shown two containers, one with a higher proportion of desirable chips than the other. Children were asked to point to the container they wanted to take a chip from. Children tended to point to the container with the higher proportion of desirable chips. In this study, experimenters also administered a variation in Piaget and Inhelder's (1975) task and found again that children did not make accurate probability judgments. However, they also found that if children completed the described task prior to the Piagetian task, they reliably made accurate predictions on the Piagetian task.

Given the mixed results in these previous studies, the present experiments include baseline control conditions that explore 3- to 7-olds' ability to make accurate and explicit verbal probability judgments after viewing a distribution. These control conditions are similar to the classic Piagetian task but were designed to be simpler and more straightforward for children.

Introduction to Experiments 1–5

This article reports findings from five experiments exploring the effects of desirability and probability on 3- to 10-year-old children's predictions. We included children from Peru as well as the United States. We also included lower-income as well as middle-income U.S. preschool children. Recently, psychologists have become conscious of the limitations of only sampling from western, educated, industrialized, rich, and democratic demographics. For example, a survey of published literature found that < 7% of published developmental psychology studies sampled children from Africa, Central and South America, Asia, Israel, and the Middle East, whereas < 1% sampled children from South or Central America (Nielsen, Haun, Kärtner, & Legare, 2017). This general lack of diversity makes it difficult to build a comprehensive picture of how development unfolds universally. Moreover, it seems plausible that cultural and socioeconomic status (SES) differences might affect the development of wishful thinking and optimism more generally, though specific comparisons and predictions are not clear given the paucity of evidence. This gap can only be addressed by actually conducting studies in a wider range of cultures and socioeconomic settings and using these findings to generate theoretical predictions; this is our strategy in the present research.

In Experiment 1, children viewed a card deck composed of two types of cards. In Experiments 2, 3, 4, and 5 children viewed a bag of plastic eggs composed of two colors. Children were asked to guess what card type or egg color had been randomly selected. The distribution was heavily skewed so that 80% of objects were of one type, and only 20% were of the other. Baseline control conditions measured children's probability judgments. In experimental conditions, the improbable outcome was also desirable. If young children can make accurate probability judgments, they should reliably predict the more likely outcome in the control conditions. If desirability alters expectancies, children should predict the unlikely (but desirable) outcome more often in the experimental conditions than in the control conditions.

Experiment 1 Methods

Participants

In the United States, participants were recruited and tested at children's science museums in the San Francisco Bay Area. The sample was predominantly middle- and upper-middle class, primarily composed of Asian (35%), Caucasian (33%), and Hispanic or Latino (17%) children. In Peru, children were recruited and tested in Innova schools located in and around Lima, Peru. This is a chain of private schools designed to serve largely lower-middle class children in Peru. Children were primarily second or third generation internal immigrants from the Peruvian highlands. Children were from an emerging middle-class background families who have traditionally been in the lower class but recently have accumulated some expendable income. All schools were located in low-income, and largely high-crime neighborhoods.

Two-hundred and sixty children participated in Experiment 1. The experimental condition included 41 U.S. 4-year-olds, 41 U.S. 6-year-olds, 23 Peruvian 4-year-olds, and 25 Peruvian 6-year-olds. The control condition included 41 U.S. 4-year-olds, 41 U.S. 6-year-olds, 23 Peruvian 4-year-olds, and 25 Peruvian 6-year-olds. See Table 1 for mean ages and age ranges. In addition, one child was dropped because of parental interference and two because of experimenter error.

Stimuli and Protocol

Experimenters used white index cards with shapes pictured on one side. The cards featured black squares and circles (United States) or triangles and circles (Peru). This study was developed in the United States then extended to children in Peru. During the initial piloting in Peru, several of the younger children called the square a triangle, so

Table 1
Summary of Participants Ages and Responses

Group	*n*	M_{age} (*SD*)	Age range	% (#) Guessed majority	95% CI for the mean	Binomial tests (*p*)
Experiment 1						
U.S. Experimental 4s	41	4.47 (0.3)	3.86–5	44 (18)	.28, .6	.533
U.S. Experimental 6s	41	6.44 (0.31)	5.94–7	54 (22)	.38, .7	.755
U.S. Control 4s	41	4.47 (0.32)	3.9–5.09	68 (28)	.53, .83	.028
U.S. Control 6s	41	6.47 (0.33)	5.95–7	88 (36)	.77, .98	< .0001
Peru Experimental 4s	23	4.54 (0.27)	3.97–4.85	35 (8)	.14, .56	.21
Peru Experimental 6s	25	6.55 (0.25)	6.06–7.0	52 (13)	.31, .73	1
Peru Control 4s	23	4.3 (0.32)	3.87–4.98	61 (14)	.39, .82	.405
Peru Control 6s	25	6.55 (0.33)	5.94–6.98	84 (21)	.69, .99	< .001
Experiment 2						
U.S. experimental	41	5.08 (1.1)	3.53–6.96	20 (8)	.07, .32	< .001
U.S. unmotivated	40	4.98 (1.05)	3.51–6.99	78 (31)	.64, .91	< .001
U.S. motivated	40	5.05 (1.04)	3.55–6.95	68 (27)	.52, .83	.039
Peru experimental	41	5.13 (1.11)	3.64–7	27 (11)	.13, .41	.004
Peru unmotivated	41	5.41 (1.13)	3.7–7.1	63 (26)	.48, .79	.117
Peru motivated	46	4.82 (0.92)	3.65–6.83	70 (32)	.56, .83	.011
Experiment 3						
Low-SES experimental	20	4.47 (0.59)	3.5–5.46	15 (3)	−.02, .32	.003
Low-SES control	25	4.46 (0.59)	3.43–5.59	60 (15)	.39, .81	.424
Experiment 4						
U.S. experimental	40	6.23 (0.89)	4.94–7.86	43 (17)	.26, .59	.43
U.S. control	40	6.23 (0.88)	4.98–7.84	78 (31)	.64, .91	< .001
Peru experimental	40	6.22 (0.94)	5.01–7.94	25 (10)	.11, .39	.002
Peru control	40	6.36 (0.88)	5.00–7.95	80 (32)	.67, .93	< .001
Experiment 5						
U.S. experimental	32	8.69 (1.04)	7.02–10.55	72 (23)	.55, .88	.02
Additional data						
Peru experimental	16	7.8 (0.31)	7.05–8.27	31 (5)	.06, .57	.21

Table includes subject numbers, mean ages (1 *SD* of the mean age), age ranges, the percent (and number) of participants who guessed the majority card or egg, 95% CIs for the mean number participants who guessed the majority card or egg, and *p*-values from two-tailed binomial tests comparing the pattern of responses to chance. SES = socioeconomic status.

experimenters replaced the square shape with a triangle shape. The experimenters also used small bins filled with colored plastic containers. The containers held prizes. U.S. participants were tested in English, and Peruvian participants in Spanish. Study protocols were translated and back-translated by bilingual research personnel.

Procedure

In the United States, children were tested in a quiet corner of the museum. In Peru, children were tested in private office spaces in their schools. First, the experimenter asked children if they liked prizes. Upon affirmation, the experimenter told children they could win prizes. Children were instructed to select one container from a bin and were told that it had a prize inside. Before the child could open the container, the experimenter placed it to the side of the table, explaining that the child might be able to win the prize later.

The experimenter next introduced participants to a deck of 20 cards and told them that the cards had circles and squares (United States.) or circles and triangles (Peru) on them. The experimenter explained that they were going to mix the cards up, then randomly select one card from the deck.

Next, the experimenter explained the prize contingencies, which differed across conditions. In the *control* condition, participants were told that they would win an additional prize, regardless of the experimenter's card selection from the deck. In the *experimental* condition, participants were told that they would only win an additional prize if one of the types of cards (i.e., the unlikely card) was selected, and would lose their initial prize if the other type was selected. Thus, in the control condition, the children believed they would receive two prizes regardless of the experimenter's selection, whereas in the experimental condition they believed they would receive two prizes if the experimenter selected the unlikely card, and no prizes if the experimenter selected the likely card.

In the experimental condition, children were asked to state which of the card types they wanted. If they said they wanted the card that resulted in no prizes, the experimenter explained the prize contingencies again, and asked the question again. All but one child agreed they wanted the experimenter to select the desirable card.

Next, the experimenter sorted all cards face up by shape type. Then, the experimenter and participants counted the number of cards of each shape. Card decks contained 16 cards of the majority shape, and four cards of the minority shape. In the experimental condition, the majority card was associated with loss, whereas the minority card was associated with gain.

Following this, the experimenter turned the cards over, mixed them up, selected one card randomly from the deck, and placed it face down on the table. Children were asked to guess which card the experimenter had selected (e.g., "What card do you think this is?"). A memory check was introduced part way through data collection. After making a prediction, 219 children were also asked to state the majority card (e.g., "Do you remember which card there was more of?"). Majority card type was counterbalanced.

Experiment 1 Results

Children were scored on whether they stated that the majority card type had been selected. A binary logistic regression explored if children predicted the majority card type using condition, country, and age group (categorical: 4 vs. 6) as predictor variables. The resulting model was statistically significant, $\chi^2(3) = 33.129$, $p < .0001$, Nagelkerke $R^2 = .163$; there was a main effect of condition, $\chi^2 = 23.127$, $df = 1$, $p < .0001$, and age, $\chi^2 = 8.266$, $df = 1$, $p = .004$, but not of country, $p = .355$, *ns*. Overall, children were more likely to choose the majority card type in the control condition than they were in the experimental condition, and older children chose the majority card type more often than younger children (see Figure 1).

Next, we were curious if developmental differences were specific to either the control or experimental condition. To explore this, we split the participants by condition, and used two binary logistic regressions to explore if age impacted predictions. For the control condition, the model was significant, $\chi^2(1) = 7.875$, $p = .005$, Nagelkerke $R^2 = .088$; and age was a significant predictor variable, $\chi^2 = 7.266$, $df = 1$, $p = .007$. Age, however, did not impact performance in the experimental condition, the model was not significant, $\chi^2(1) = 2.013$, $p = .156$, Nagelkerke $R^2 = .021$; and age was not a significant predictor variable, $\chi^2 = 1.997$, $df = 1$, $p < .158$. In sum, with age, children provided more accurate responses in the control condition, however, 4- and 6-year-olds were equally likely to engage in wishful thinking in the experimental condition.

In the control condition, children reliably predicted the majority card type; 99 of 130 children (or 76%; $SD = 0.43$; 95% CI [.69, .84]) predicted the majority card type, which is significantly greater than chance, $p < .0001$, two-tailed binomial test. This was also true when both the 4- and 6-year-old age groups were considered separately (4-year-olds: 42 of 64, or 66%; $SD = 0.48$, 95% CI [.54, .78]; $p = .017$, two-tailed; 6-year-olds: 57 of 66, or 86%; $SD = 0.35$, 95% CI [.78, .95]; $p < .0001$, two-tailed). In the experimental condition, 61 of 130 children (or 47%; $SD = 0.5$; 95% CI [.38, .56]) chose the majority card, which is not significantly different from chance, $p = .539$, *ns*. A power analysis was conducted using the program G*power Version 3.1.9.3., and the means presented earlier. Results suggest a total sample size of 90 to find the main effect of condition (with power $1 - \beta$ set to .80; $\alpha = .05$; two-tailed test), indicating that the sample size in this study was more than adequate.

Two-hundred and nineteen children were asked if they remembered which card there was more of, as well as which card there was less of. In the *control* condition, 83% of children answered both

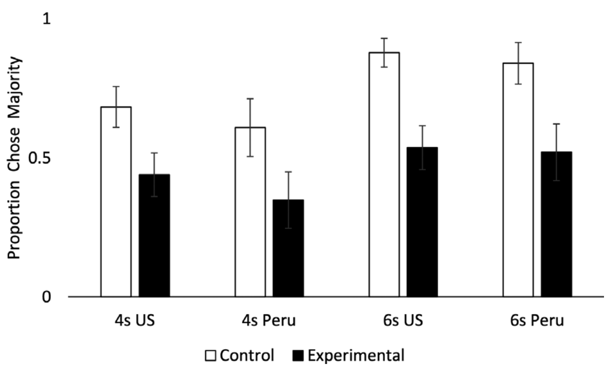

Figure 1. Proportion of children who stated the majority response in Experiment 1. Error bars represent 1 *SEM*.

questions correctly. In the *experimental* condition, 89% of children correctly answered both questions (chance is 25%). Looking only at children who responded correctly to the memory checks, in the control condition 62 of 90 children guessed the majority, and in the experimental condition 43 of 98 did so. A Fisher's exact test confirms that the difference between conditions remained significant, $p = .0007$. Children's optimism in the experimental condition cannot be explained by a failure to remember the distribution.

Audible video recordings were obtained for 72 children in the control condition, and 85 in the experimental condition. Using these recordings, children were retroactively scored on whether they correctly stated the prize contingencies associated with each of the cards without prompting, and without any re-explanation from the experimenter. Eighty-three percent (or 60 of 72) of the children in the control condition correctly stated that both cards would result in two prizes, whereas 65% (or 55 of 85) of children accurately recalled the prize contingencies in the experimental condition. Looking only at these children, 75% (or 45 of 60) stated the majority card in the control condition, whereas 55% (or 30 of 55) did so in the experimental condition. A Fisher's exact test reveals that the difference between conditions remained significant, $p = .031$ (two-tailed).

Experiment 1 Discussion

Study 1 suggests that 4- and 6-year-old children's verbal predictions were influenced by both desirability and probability. Children scored significantly above chance in the control condition, and there was a significant difference between control and experimental conditions, indicating an effect of wishful thinking on children's judgments.

Older children overall were more likely to state the majority card than younger children in the control condition, however, there was no effect of age in the experimental condition. This indicates that both 4-and 6-year-olds were equally likely to engage in wishful thinking, even though 6-year-olds displayed more advanced probabilistic reasoning skills. Four-year-olds did score above chance in the control condition, however, their performance was still not impressive.

One concern is that children scored at chance in the experimental condition. It is possible that desirability biased children's answers, resulting in a pattern of responses that was meaningfully different from the control condition, but coincidently at chance. Alternatively, it is possible that tracking the two levels of prize contingencies (0 vs. 2 prizes), as well as both gain and loss contingencies was difficult for children, and children resorted to guessing.

This raises questions about whether the experimental design was suitable for young children. For example, gathering and shuffling the cards took a while, and there was a substantial gap in time between when children viewed the distribution, and when the card was selected. The memory checks at the beginning of the experiment were lengthy, and many children seemed to lose interest during these. Given this, it is possible that the developmental trends were caused, at least in part, by developmental differences in working memory, or attentional regulation. There was also no reward for correct answers in the control; this may have influenced younger children's responses.

In Experiment 2, we attempted to control for these possibilities and explore whether we could replicate the previous experiment using different materials and procedure. First, we used a shorter, more visually simple version of this task. The experimenter chose a random sample from a collection of objects which visibly included more of one type than another (an "urn" type of probability task). Both infants, implicitly, and older preschool children, explicitly, have demonstrated that they understand probability in "urn" tasks and assume random sampling (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Denison & Xu, 2014; Xu & Garcia, 2008). We also included an additional control condition, where children were incentivized for providing a correct answer, to explore the possibility that greater motivation might improve the younger children's performance on the control task. In addition, this new control condition required children to track three colors of eggs that contained two different amounts of prizes (0 or 2), matching the cognitive demands of the experimental condition.

Experiment 2 Methods

Participants

One-hundred and twenty-one North American and 128 Peruvian 3- to 6-year-olds participated. Children were divided across three conditions: the experimental condition, the motivated control condition, and the unmotivated control condition. See Table 1 for further information on subject numbers and age. Additionally, three participants were tested and not included: one child voluntarily withdrew, one child failed to provide a response, and one was dropped due to experimenter error.

Stimuli

In Peru, the experimenter used a special blue plastic egg (motivated control condition only), a white cloth, a brown paper bag, and a clear plastic bag containing 10 yellow and purple plastic eggs. Some eggs contained stickers. In the United States, stimuli were similar, but egg colors differed. The experimenter used a silver plastic egg (motivated control condition only), and a clear plastic bag containing 10 yellow and blue eggs. Experiment 2 was initially developed and tested in Lima, Peru, where experimenters had limited access to study stimuli. U.S. experimenters changed the egg color from purple to blue because it was thought that some children (mostly girls) might have a strong preference for purple eggs, and this might impact the results. The special blue egg (motivated control condition) was painted silver in the United States to differentiate it from the other eggs.

Procedure

Experimental Condition

Children were first introduced to a clear bag containing purple and yellow (Peru) or blue and yellow (United States) plastic eggs. The color distribution was 8 to 2, and the majority color was counterbalanced. To ensure that children took note of the different colors and could differentiate them, children were asked to point to one of each color of egg. Then, the experimenter told participants that the minority egg color contained two stickers and the majority color did not have any stickers. Following this, the experimenter took one of each type of egg out of the bag, opened them up, and showed the children what was inside. The example eggs were then reassembled and placed back inside of the plastic bag. Next, the experimenter asked participants to point to an egg containing two stickers and an egg containing no stickers. The experimenter and child counted out loud the number of each type. Then the experimenter again asked the child if they remembered which egg had two stickers, and which egg had no stickers. Next the experimenter held the clear plastic bag of eggs over a brown paper bag and explained that they were going to place the clear bag into the paper bag and select one egg without looking into the bag, and the child would have to guess the color. The experimenter also told participants that if the egg had prizes inside, the child could keep them.

Then, the experimenter lowered the clear bag into the opaque bag, reached in and placed a white cloth over a randomly selected egg. The selected egg was immediately placed on the table, still under the cloth and covered by the experimenter's hands. The experimenter said, "*Hmm*, I wonder what color it is. What color do you think it is? Purple or yellow?" The order in which the two colors were listed was counterbalanced.

Unmotivated Control Condition

This condition was identical to the experimental condition, except that all the eggs contained two stickers.

Motivated Control Condition

First, the experimenter showed participants a special blue egg (Peru) or silver egg (United States), explaining that it contained two stickers, which could be won. They then opened the special egg to show that it actually contained two stickers.

The rest of the procedure was similar to the other conditions, except for two differences. First, there were no stickers inside any of the other eggs (i.e., the eggs in the clear plastic bag that formed the distribution from which the experimenter was sampling). Second, children were told that they would win the stickers inside the special egg if they correctly guessed what color the experimenter selected from the bag. This is different from the experimental and unmotivated control conditions, in which children were told that they would win whatever was inside of the egg *selected* from the bag. This condition was included to test whether children who were motivated to be accurate in their predictions would perform better than children who were not. It also better matched the cognitive demands of the experimental condition, in that children had to track two amounts of prizes (0 vs. 2 stickers) across three colors of eggs, rather than just one type of prize contingency (two stickers) across two colors of eggs.

In all three conditions, after children guessed what color of egg was under the cloth, they were asked to recall the egg color there was "more of." In the experimental condition, they were also asked to recall which egg color they wanted. We included this question *after* the child's guess in Experiment 2—as opposed to *before* the child's guess in Experiment 1—to control for the possibility that stating a preference might have primed participants' guesses. At the end of the procedure, the experimenter revealed the egg color. All children were immediately given prizes, regardless of the outcome- either the prizes inside of the egg, or a reward for playing the game.

Experiment 2 Results

Children were scored on whether they guessed that the majority egg color had been drawn from the bag. First a binary logistic regression was used to compare the experimental to the unmotivated control condition, using country, exact age (as a continuous variable, given that the children's ages ranged from 3 to 6), and condition as predictor variables. The resulting model was statistically significant, $\chi^2(3) = 46.134$, $p < .0001$, Nagelkerke $R^2 = .329$. Analyses revealed a main effect of condition, $\chi^2 = 32.970$, $df = 1$, $p < .0001$, and age, $\chi^2 = 7.43$, $df = 1$, $p = .006$, but not country, $p = .419$, *ns*. As in Experiment 1 children chose the majority egg more often in the unmotivated control condition than in the experimental condition, and older children chose the majority egg color more often than younger children.

A second binary logistic regression compared the motivated control condition to the experimental condition. Country, exact age (as a continuous variable), and condition were entered into the model as predictor variables. The resulting model was also statistically significant, $\chi^2(3) = 39.927$, $p < .0001$, Nagelkerke $R^2 = .283$. Analyses again revealed a main effect of condition, $\chi^2 = 32.71$, $df = 1$, $p < .0001$. Age trended toward being a significant predictor, $\chi^2 = 3.151$, $df = 1$, $p = .076$. Country was not significant, $p = .444$, *ns*. Children chose the majority egg more often in the motivated control condition than in the experimental condition, and older children trended toward choosing the majority egg color more often than younger children. As in Experiment 1, children demonstrated a wishful thinking bias.

Next, we split data by condition, and used three binary logistic regressions to explore if age influenced responses within each condition individually. In the experimental condition, age did not predict children's performance. The model was not significant, $\chi^2(1) = 2.528$, $p = .112$, Nagelkerke $R^2 = .046$, and age was not a significant predictor, $\chi^2 = 2.457$, $df = 1$, $p = .117$. Age, however, did impact children's performance in the unmotivated control condition. The model was significant, $\chi^2(1) = 5.132$, $p = .023$, Nagelkerke $R^2 = .087$, and age was a significant predictor variable, $\chi^2 = 4.664$, $df = 1$, $p = .031$. In the motivated control condition, the model was again not significant, $\chi^2(1) = .806$, $p = .369$, Nagelkerke $R^2 = .013$, and age was not a significant predictor variable, $\chi^2 = .79$, $df = 1$, $p = .374$. In sum, with age, children provided more accurate responses in the unmotivated control condition, however, age did not impact responses in either the experimental or motivated control conditions.

Two-tailed binomial tests confirmed that children chose the majority option significantly above chance in the unmotivated control condition (57 of 81, or 70%; $SD = 0.46$; 95% CI [.6, .81]), $p = .0003$, as well as in the motivated control condition (59 of 86, or 69%; $SD = 0.47$; 95% CI [.59, .79]), $p = .0007$; there were no differences between the two control conditions, $p = .867$, *ns*, two-tailed Fisher's exact test. In contrast, children scored significantly *below* chance in the experimental condition. Only 19 of 82, or 23% ($SD = 0.42$; 95% CI [.14, .33]) of children stated the experimenter had selected the majority egg, $p < .0001$. Most children believed that the experimenter had selected the desirable, yet highly improbable, egg. A power analysis was conducted using the software G*power, and the means presented earlier. Results suggest a total sample size of 37 to find the difference between the experimental and unmotivated control conditions, and a total sample size of 39 to find the difference between the experimental and motivated control conditions (with power $1 - \beta$ set to .80; $\alpha = .05$; two-tailed test). Again, this suggests that the sample size used in this study was more than adequate and validates the sample sizes used in the following experiments, which enlist a similar paradigm.

In the experimental condition, 82% of participants stated they wanted the egg with the prizes, and 76% correctly stated which egg there was more of. In the motivated control condition, 85% correctly stated the majority egg color, and 91% did so in the unmotivated control condition. Looking only at children who passed the memory check questions, and also stated they wanted the egg with the prizes, in the unmotivated control 50 of 74 (68%) children predicted the majority egg, in the motivated control condition 48 of 72 (67%) guessed the majority egg, and in the experimental condition 11 of 51 (22%) predicted the majority egg. Fisher's exact tests confirm that the difference between the unmotivated control and experimental conditions remained significant, $p < .0001$, as did the difference between the motivated control and experimental conditions, $p < .0001$.

Experiment 2 Discussion

Like Experiment 1, Experiment 2 indicated a difference between the experimental and control conditions, supporting the hypothesis that young children engage in wishful thinking. In the experimental condition, very few children, only 23%, predicted the likely outcome, which was significantly below chance. Age did not impact children's responses in the experimental condition.

In control conditions, children again made accurate probability judgments, scoring above chance on both control conditions. Children's performance in the unmotivated control condition was generally similar to their performance in Experiment 1, where children's accuracy increased with age. Performance in the motivated control condition, however, did not show an age effect. This suggests that the age differences in control conditions might reflect motivational differences. Children's mean scores, however, were similar across all control conditions.

In the experimental condition in Experiment 2, most children (77%) stated the unlikely (and desirable) outcome, whereas in Experiment 1, only 53% did so. This may be because the design in Experiment 2 was simpler and more straightforward for children, decreasing the noise in children's responses. A few changes in particular may have made the experimental condition easier for young children to follow. First, prizes were inside of the eggs, rather than contingently given to children from an external source. Second, in Experiment 2, the eggs simply had two prizes or no prizes. In Experiment 1 children were given an initial prize, then, based on the experimenter's selection they were either given 1 more prize, or the initial prize was taken away. This may have been confusing. Additionally, Experiment 2 was faster, and more visually appealing, possibly making it easier for children to track the information.

Experiment 3

In Experiment 3, we extend this paradigm to 3- to 5-year-old children enrolled in Head Start programs in Berkeley, California. To be eligible for enrollment in Head Start, families' income must fall below the federal poverty level, which, at the time of testing, was below $24,600 for a family of 4 ("2017 Poverty Guidelines," 2017). Economists of happiness have reported that levels of optimism, happiness and life satisfaction vary by income, with people from lower SES backgrounds consistently scoring lower on these measures than those from middle- and upper-middle class backgrounds (e.g., Graham, 2017; Kahneman & Deaton, 2010). This could indicate that lower-SES children may be less prone to a wishful thinking bias (as it is a type of optimism). However, Marks (1951) found that SES did not impact wishful thinking in grade school children, suggesting that the lower-SES U.S. children might score similarly to the samples previously tested.

Experiment 3 Methods

Participants

Experiment 3 included 45 children. Twenty children participated in the experimental condition (M_{age} = 4.47, *SD* = 0.59; range = 3.5–5.46), and 25 in the control condition (M_{age} = 4.46, *SD* = 0.6; range = 3.43–5.59). Children were recruited and tested at Head Start programs in Berkeley, CA.

Methods

Methods were identical to the experimental and unmotivated control conditions of Experiment 2; all children viewed a clear bag of blue and yellow eggs. Children were tested in a quiet room or hallway at their preschool.

Results

Children were scored according to whether they guessed the majority egg color. A binary logistic regression measured if age (as a continuous variable) and condition predicted majority response. The model was statistically significant $\chi^2(2) = 12.416$, $p = .002$, Nagelkerke $R^2 = .326$. Analyses revealed a main effect of condition, $\chi^2 = 8.339$, $df = 1$, $p = .004$. There was no effect of age, $p = .135$, *ns*.

Two-tailed binomial tests compared responses to chance. In the experimental condition, only 3 of 20 (or 15%; *SD* = 0.37; 95% CI [−.02, .32]) children guessed the majority egg, which is significantly below chance, $p = .003$; in the control condition 15 of 25 (or 60%; *SD* = 0.5; 95% CI [.39, .81]) children guessed the majority egg, which is not significantly different from chance, $p = .424$, *ns*.

In the experimental condition, all but one child (95%) stated they wanted the desirable egg. In the control condition, 68% of children correctly stated the majority egg color, whereas 60% of children did so in the experimental condition. Looking only at these children, 9 of 17 (53%) children stated the majority color in the control condition, and only 1 of 11 (9%) did so in the experimental condition. The difference across conditions remained significant, p = .041 (two-tailed Fisher's exact test).

Experiment 3 Discussion

Experiment 3 extends findings from Experiments 1 and 2 to lower income children in the United States. Three-to 5-year-old children enrolled in Head Start programs displayed very high levels of wishful thinking, where 85% of children provided an optimistically biased response in the experimental condition. These responses are similar to the middle-income U.S. American, and Peruvian 3- to 5- year-olds.

Experiment 4

Age did not influence 4- to- 6-year-olds' responses in the experimental conditions of Experiments 1, 2 and 3. Children displayed high levels of wishful thinking across experiments; for example, 77% of children stated that the highly improbable, yet desirable egg was selected in Experiment 2. Intuitively, it seems that adults would not show such a strong bias, and that with age, this bias should attenuate, at least to some extent. We explore this more in Experiments 4 and 5 by extending this paradigm to older children.

Experiment 4 Methods

Participants

Eighty U.S. and 80 Peruvian 5- to 7-year-olds participated in this study. Participant demographic information and testing setup were similar to that of Experiments 1 and 2. See Table 1 for more information on subject ages. Additionally, one 6-year-old was tested and not included in the final sample due to experimenter error.

Methods

Procedures were identical to the experimental and unmotivated control conditions of Experiments 2 and 3 with two exceptions. First, children were not told there were stickers inside of the eggs or shown the prizes. Rather, they were told that the eggs contained "prizes." This was to control for any developmental differences in the desirability of specific types of prizes. Second, light blue and yellow eggs were used for children from both Peru and the United States.

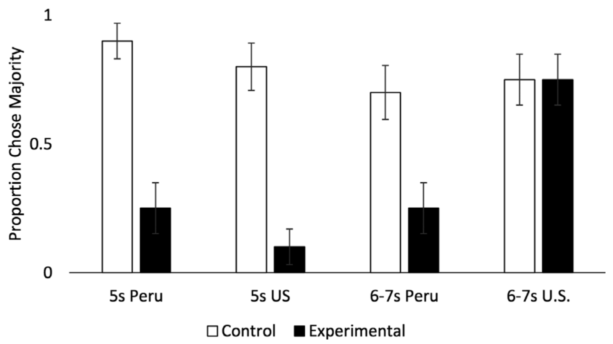

Figure 2. Proportion of children who stated the majority response in Experiment 4. Error bars represent 1 *SEM*.

Results

A binary logistic regression explored if children's responses were predicted by age (continuous) country (Peru vs. United States) and condition (experimental vs. control). The model was significant, $\chi^2(3) = 38.456$, $p < .0001$, Nagelkerke $R^2 = .286$. Analyses revealed a main effect of condition, $\chi^2 = 29.984$, $df = 1$, $p < .0001$, indicating that children were more likely to guess the majority egg color in the control condition than in the experimental condition. There was a trending effect of age, $\chi^2 = 2.98$, $df = 1$, $p = .084$ indicating that older children were slightly more likely to choose the majority egg color than younger children; country was an insignificant predictor, $p = .258$, *ns* (Figure 2).

Next, we split participants into condition, and two binary logistic regressions were used to explore if age (as a continuous variable) predicted children's responses in the experimental and control conditions individually. Age did not predict children's responses in the control condition, the model was not significant, $\chi^2(1) = 1.321$, $p = .25$, Nagelkerke $R^2 = .025$, and age was not a significant predictor, $\chi^2 = 1.322$, $df = 1$, $p = .25$, *ns*. Age was, however, a significant predictor in the experimental condition. The model was significant, $\chi^2(1) = 10.227$, $p = .001$, Nagelkerke $R^2 = .166$, and age was a significant predictor variable $\chi^2 = 9.202$, $df = 1$, $p = .002$. This finding indicates that with age, children were less likely to display wishful thinking.

Two-tailed Binomial tests were again used to compare children's responses to chance. Overall, in the control condition, 63 of 80 (or 79%; $SD = 0.41$, 95% CI [.7, .88]) children predicted the majority response, which is significantly above chance

$p < .0001$. This can be contrasted with the experimental condition, where 27 of 80 (or 34%; $SD = 0.48$, 95% CI [.23, .44]) children stated the majority response, which is significantly below chance, $p = .005$.

At the end of the experiment, 94% of children in the control condition and 81% of children in the experimental condition correctly recalled the majority color. In the experimental condition, all but seven children (91%) stated they wanted the egg with the prizes. After removing the children who answered these questions incorrectly, 58 of 75 (or 77%) children in the control condition and 16 of 61 (or 26%) children in the experimental condition guessed the majority egg. A Fisher's exact test confirmed that the difference between conditions remained significant, $p < .0001$.

Experiment 4 Discussion

Experiment 4 again found a strong effect of wishful thinking. Children were much more likely to state the majority egg color in the control condition than in the experimental condition. We also saw that age influenced children's responses in the experimental condition; older children were less likely than younger children to display a wishful thinking bias. No difference was observed in the control condition. In sum, Experiment 3 replicates findings from Experiments 1 and 2, and provides initial support for the hypothesis that wishful thinking declines with age. In Experiment 5, we follow-up on this finding by testing 7-to 10-year-old children in the United States.

Experiment 5 Methods

Participants

Participants included thirty-two 7- to 10-year-old children ($M_{age} = 8.69$, $SD = 1.04$, range = 7.02–10.55) from the San Francisco Bay Area. Additionally, one child was tested and dropped because the experimenter did not obtain a birthdate. Participant demographic information is similar to Experiments 1–4. Due to limitations in access to Peruvian children, data analysis for Experiment 5 is restricted to the United States. However, experimenters were able to collect a partial sample of Peruvian 7- and 8-year-olds ($n = 16$), and these children are included in Tables 1 and 2 as well as the meta-analysis after Experiment 5. Demographic information for these children is similar to that in the previous experiments.

Methods

Methods were identical to the experimental condition of Experiment 4. Given that the previous samples demonstrated proficient probabilistic reasoning skills, and age is the primary variable of interest, a control condition was not included.

Results

A binary logistic regression was used to explore if age (continuous variable) predicted children's responses in the experimental condition. Results indicated a significant effect of age. The model was significant, $\chi^2(1) = 12.153$, $p < .001$, Nagelkerke $R^2 = .454$, and age was a significant predictor variable, $\chi^2 = 6.747$, $df = 1$, $p = .009$.

Twenty-three of 32 (or 72%; $SD = 0.46$, 95% CI of the mean [.55, .88]) children stated the majority egg color, which is significantly greater than chance would predict, $p = .02$ (two-tailed binomial test), indicating that most 7- to 10-year-old children did *not* demonstrate wishful thinking.

At the end of the experiment, three children did not correctly state the majority egg color, and one child did not state that they wanted the egg color with the prizes. Removing these children from the sample indicates that 21 of 28 (or 75%) of children still chose the majority egg color, which is still significantly above chance, $p = .013$ (two-tailed binomial test). A regression analysis on these children revealed age trends comparable to those described earlier.

Experiment 5 Discussion

Experiment 5 extends on findings from Experiment 4 and suggests that children's tendency to engage in wishful thinking continues to decline between 7 and 10 years of age. In this study, most children did not display a bias, however, previous studies have found evidence for wishful thinking in grade school children and even adults. Notably, however, the adult research suggests that the bias is only reliably observed when the ratio of the two outcomes is near 50:50. In this study, the ratio was much more extreme, 80:20. Given this, it would be interesting to measure 7- to 10- year-olds' predictions with less extreme ratios.

Additionally, previous studies used slightly different methods than this study. For example, in this study, we visually display the ratios directly to children, whereas Marks (1951) verbally stated the ratios to children. This study also asked children to

Table 2
Summary of Participants Included in the Meta-analysis on Experimental Conditions and Responses Given

Group	n	M_{age} (SD)	Age range	% (#) Guessed majority	95% CI for the mean	Binomial tests (p)
Peru and U.S.						
All participants	210	6.29 (1.6)	3.53–10.55	35 (74)	.29, .42	< .0001
Youngest	53	4.35 (0.54)	3.53–5.11	17 (9)	.07, .27	< .0001
Second youngest	53	5.64 (0.31)	5.13–6.16	21 (11)	.09, .32	< .0001
Second oldest	52	6.78 (0.38)	6.18–7.47	37 (19)	.23, .5	.07
Oldest	52	8.42 (0.86)	7.49–10.55	67 (35)	.54, .8	.018
U.S.						
All participants	113	6.51 (1.77)	3.53–10.55	43 (48)	.33, .52	.132
Youngest	28	4.37 (0.58)	3.53–5.11	11 (3)	−.02, .23	< .0001
Second youngest	27	5.74 (0.26)	5.29–6.16	19 (5)	.03, .34	.002
Second oldest	26	6.83 (0.37)	6.18–7.47	50 (13)	.29, .71	1.155
Oldest	32	8.78 (0.92)	7.56–10.55	84 (27)	.71, .98	.0001
Peru						
All participants	97	6.02 (1.33)	3.64–8.27	27 (26)	.18, .36	< .0001
Youngest	25	4.33 (0.51)	3.64–5.07	24 (6)	.06, .42	.015
Second youngest	26	5.53 (0.32)	5.13–6.12	23 (6)	.06, .4	.009
Second oldest	26	6.73 (0.38)	6.22–7.43	23 (6)	.06, .4	.009
Oldest	20	7.85 (0.21)	7.49–8.27	40 (8)	.16, .64	.503

Table includes subject numbers, mean ages (1 *SD* of the mean age), age ranges, the percent (and number) of participants who stated the majority egg color, 95% CIs for the mean number of participants who stated the majority egg color, and *p*-values from two-tailed binomial tests comparing majority responses to chance. Data are split into quartiles based on participants' ages, then further subdivided by country; quartiles are age matched across countries, with the exception of the eldest quartile.

make judgments about an event that had already occurred (the egg was already selected when children were asked to make the judgment), whereas Marks (1951) and others asked participants to make a prediction about a card that would be selected in the near future. These types of methodological differences could have influenced participant's tendency to display a wishful thinking bias.

Meta-Analysis

To further explore the observed age trend on wishful thinking, we conducted a meta-analysis on the experimental conditions from Experiments 2, 4, and 5. This resulted in a total of 194 children. To get a more complete picture of developmental trends within each country, we also included 16 Peruvian 7- and 8-year-olds ($M_{age} = 7.8$; $SD = 0.31$; range = 7.05–8.27) who were tested for Experiment 5, however, due to limitations, a complete sample was not collected. This resulted in a total sample size of 210 children ($M_{age} = 6.29$; $SD = 1.6$; range = 3.53–10.55).

First, a binary logistic regression explored the effect of age on all children's responses. The model was significant, $\chi^2(1) = 37.973$, $p < .0001$, Nagelkerke $R^2 = .23$, and age was a significant predictor variable, $\chi^2 = 30.046$, $df = 1$, $p < .0001$.

Next, we split the data set by country. This resulted in 113 U.S. children ($M_{age} = 6.51$; $SD = 1.77$, range = 3.53–10.55) and 97 Peruvian children ($M_{age} = 6.02$; $SD = 1.33$; range = 3.64–8.27). Two binary logistic regressions explored the effect of age on responses. The model for the U.S. children was highly significant, $\chi^2(1) = 42.084$, $p < .0001$, Nagelkerke $R^2 = .418$, and age was a significant predictor variable, $\chi^2 = 25.976$, $df = 1$, $p < .0001$. In Peru, the model did not approach significance, $\chi^2(1) = 1.12$, $p = .29$, Nagelkerke $R^2 = .017$, and age was not a significant predictor variable, $\chi^2 = 1.102$, $df = 1$, $p = .294$. Additionally, we split participants into quartiles based on age, then further divided them by country. Table 2 presents an overview of ages and responses.

General Discussion

Across all experiments, 3- to 5-year-old children reliably displayed a very strong wishful thinking bias. This finding held across cultures and socioeconomic backgrounds. In the United States, this bias gradually declined with age, with preschool children displaying the strongest bias and 7- to 10-year-olds displaying no bias. The meta-analysis suggests that this trend may be different in Peru, possibly weaker, or later developing, however, further research should be conducted with older Peruvian children given that the age ranges were slightly different across populations.

Across experiments, children made accurate probabilistic judgments in control conditions, with the exception of the lower income children in Experiment 3. In Experiments 1 and 2, older children made more accurate judgments in the control conditions than younger children. Even so, the youngest children still reliably provided accurate responses. These results confirm the earlier findings on probabilistic reasoning, suggesting that children's success may be largely dependent on the task demands of the experimental design. With appropriately simplified materials, children under seven can give explicit and accurate probabilistic judgments.

We found that age influenced children's judgments in the experimental condition; older children were less likely to display wishful thinking than younger children. Previously, researchers have measured wishful thinking in school age children and adults, and those results suggest that wishful thinking may continue to decline during development. This raises questions about what specifically changes with age, and why older children and adults are less likely to engage in wishful thinking than young children.

One possibility is that there is simply a strong early, perhaps even inbuilt, tendency for desires to causally influence predictions and that this tendency becomes weaker with age. However, these results could also suggest that young children's beliefs about uncertain outcomes undergo conceptual revision over development, and these beliefs, rather than a direct influence of desires on predictions, are responsible for changes in wishful thinking. Young children may initially use their desires to predict uncertain outcomes, or even believe that their desires have a causal impact on outcomes. Indeed, in some cases this may be a reasonable assumption, for example, in cases where people can actually exert control over outcomes. In addition, adults often modify outcomes to be consistent with infants and young children's desires; they help children get what they want, providing further support for this belief. As children get older, they may encounter more situations where they do not get what they want and where the link between desires and outcomes is more tenuous. Over the course of time, children may begin to realize that desires do not always lead to outcomes, and instead rely on other information to make predictions, such as the likelihood evidence in the present studies.

Alternatively, people may continue to have a strong disposition toward wishful thinking throughout development, either intrinsically, or as a result of beliefs, but their other beliefs about randomness and probability could undergo conceptual change with development and offset this tendency. As their understanding of probability improves, children may begin to override the tendency to engage in wishful thinking. Of course, changes could also occur along several dimensions simultaneously.

These competing hypotheses can all explain why adults and older children still show some evidence of a wishful thinking bias. Adults and older children could simply hold a weaker desire bias, or a weaker belief that desires cause outcomes, resulting in less biased inferences in both cases. If so, adults and older children should be less likely than younger children to display wishful thinking across a variety of situations. Alternatively, adults and older children may develop a stronger belief in the alternative hypothesis that probability influences the outcome. If so, wishful thinking might reemerge when evidence in favor of the alternative hypothesis is weak or nonexistent (e.g., ratios are less extreme, or no probability evidence is provided), when the causal pathway toward an outcome is more convoluted and mysterious, or when participants are asked to make judgments in domains where they have limited prior knowledge.

In addition to explaining why wishful thinking changes during development, these accounts can be used to make predictions about how wishful thinking relates to childhood optimism more generally. If wishful thinking is generally responsible for optimism and children's wishful thinking declines, then we should see optimism decline at a similar pace across domains. Moreover, it should be possible to explore whether there are correlations between changes in wishful thinking and in other kinds of optimism. However, if children display less wishful thinking because they develop a stronger belief in a competing hypothesis, then developmental changes in optimism that result from wishful thinking should differ across domains, and should depend on the availability of evidence in favor of the alternative hypothesis.

Another possibility is that reasoning about probability together with desirability could require advances in some other aspect of cognition, such as inhibitory control. Perhaps stating the probable outcome in the experimental condition requires participants to first inhibit themselves from stating the desirable outcome. If so, older children and adults may display a stronger bias under certain conditions, for example, when they are asked to make rapid judgments.

There are some limitations to the current studies. Earlier studies and the performance in the control conditions suggest that children do indeed infer a random sampling process. Moreover, the experimenters in all the studies emphasized the random nature of the events—shuffling the cards and mixing up the eggs in an opaque bag, events that even infants interpret as random processes (e.g., Denison & Xu, 2014), closing their eyes and looking away while selecting an egg, and explicitly stating that they did not know the outcome. However, it is possible that children may have thought that the experimenter intentionally “fixed” the process in a deceptive way to give them the prizes, analogous perhaps to adults intentionally letting children win card games. Given this possibility, one next step could be to explore if the findings replicate in a condition where the random process does not involve an agent.

Another possibility is that children could have stated the desirable response partly because there was no cost associated with being incorrect. The motivated control condition in Experiment 2 did suggest that motivating younger children to be correct increased their accuracy, but no conditions explored whether associating a cost or benefit with accuracy would alter children’s predictions in the experimental conditions.

In any case, these studies lend support to the hypothesis that young children, from all the backgrounds we tested, have a strong wishful thinking bias, and that wishful thinking declines with age. However, it is not yet clear exactly why young children engage in wishful thinking and what causes developmental change. Furthermore, it is not yet entirely clear how wishful thinking is related to previous developmental findings on optimism, positivity, confidence and theory of mind. Future research should more thoroughly explore these questions.

References

2017 Poverty Guidelines. (2017, January 31). *Office of the Assistant Secretary for planning and evaluation*. Retrieved from https://aspe.hhs.gov/poverty-guidelines

Boseovski, J. J. (2012). Trust in testimony about strangers: Young children prefer reliable informants who make positive attributions. *Journal of Experimental Child Psychology*, *111*, 543–551. https://doi.org/10.1016/j.jecp.2011.10.008

Boseovski, J. J., & Lee, K. (2008). Seeing the world through rose-colored glasses? Neglect of consensus information in young children’s personality judgments. *Social Development*, *17*, 399–416. https://doi.org/10.1111/j.1467-9507.2007.00431.x

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. (2013). Rational variability in children’s causal inferences: The sampling hypothesis. *Cognition*, *126*, 285–300. https://doi.org/10.1016/j.cognition.2012.10.010

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*, 335–347. https://doi.org/10.1016/j.cognition.2013.12.001

Diesendruck, G., & Lindenbaum, T. (2009). Self-protective optimism: Children’s biased beliefs about the stability of traits. *Social Development*, *18*, 946–961. https://doi.org/10.1111/j.1467-9507.2008.00494.x

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., & Slaughter, V. (1991). Young children’s understanding of changes in their mental states. *Child Development*, *62*, 98–110. https://doi.org/10.2307/1130707

Graham, C. (2017). *Happiness for all? Unequal hopes and lives in the pursuit of the American Dream*. Princeton, NJ: Princeton University Press. https://doi.org/10.1111/1475-4932.12389

Granberg, D., & Brent, E. (1983). When prophecy bends: The preference-expectation link in U.S. presidential elections, 1952–1980. *Journal of Personality and Social Psychology*, *45*, 477–491. https://doi.org/10.1037/0022-3514.45.3.477

Hayes, S. P. (1936). The predictive ability of voters. *The Journal of Social Psychology*, *7*, 183–191. https://doi.org/10.1080/00224545.1936.9921660

Heyman, G. D., & Giles, J. W. (2004). Valence effects of reasoning about evaluative traits. *Merrill-Palmer Quarterly*, *50*, 86–109. https://doi.org/10.1353/mpq.2004.0004

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality*, *21*, 329–335. https://doi.org/10.1111/j.1467-6494.1953.tb01775.x

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 16489–16493. https://doi.org/10.1073/pnas.1011492107

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*, 95–121. https://doi.org/10.1037/0033-2909.133.1.95

Krizan, Z., & Windschitl, P. D. (2009). Wishful thinking about the future: Does desire impact optimism? *Social and Personality Psychology Compass*, *3*, 1–17. https://doi.org/10.1111/j.1751-9004.2009.00169.x

Lockhart, K. L., Chang, B., & Story, T. (2002). Young children’s beliefs about the stability of traits: Protective optimism? *Child Development*, *73*, 1408–1430. https://doi.org/10.1111/1467-8624.00480

Lockhart, K. L., Goddu, M. K., & Keil, F. C. (2017). Overoptimism about future knowledge: Early arrogance? *The Journal of Positive Psychology*, *12*, 36–46. https://doi.org/10.1080/17439760.2016.1167939

Lockhart, K. L., Nakashima, N., Inagaki, K., & Keil, F. C. (2008). From ugly duckling to swan? Japanese and American beliefs about the stability and origins of traits. *Cognitive Development*, *23*, 155–179. https://doi.org/10.1016/j.cogdev.2007.08.001

Marks, R. W. (1951). The effect of probability, desirability, and "privilege" on the stated expectations of children. *Journal of Personality*, *19*, 332–351. https://doi.org/10.1111/j.1467-6494.1951.tb01107.x

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and children's theory of mind. *Cognitive Development*, *10*, 467–482. https://doi.org/10.1016/0885-2014(95),90023-3

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

Ogburn, W. F. (1934). Studies in prediction and the distortion of reality. *Social Forces*, *13*, 224–229. https://doi.org/10.2307/2570338

Parsons, J. E., & Ruble, D. N. (1977). The development of achievement related expectancies. *Child Development*, *48*, 1075–1079. https://doi.org/10.2307/1128364

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London, UK: Routledge and Kegan Paul Ltd.

Plumert, J. M. (1995). Relations between children's overestimation of their physical abilities and accident proneness. *Developmental Psychology*, *31*, 866–876. https://doi.org/10.1037/0012-1649.31.5.866

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

Stipek, D. J., & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment? *Journal of Educational Psychology*, *80*, 352–356. https://doi.org/10.1037/0022-0663.80.3.352

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*, 521–538. https://doi.org/10.2307/1130719

Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development*, *55*, 1983–1989. https://doi.org/10.2307/1129773

Xu, F., & Garcia, V. (2008). Intuitive stastics by 8-month-old infants. *PNAS*, *105*, 5012–5015. https://doi.org/10.1073/pnas.0704450105.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development*, *33*, 769–780. https://doi.org/10.12691/education-3-4-21