ijbd

International Journal of Behavioral Development
1–9

DOI: 10.1177/0165025420905344
journals.sagepub.com/home/jbd
SAGE

Preschoolers in Belarus and Turkey accept an adult's counterintuitive claim and do not spontaneously seek evidence to test that claim

Samuel Ronfard,[1] Burcu Ünlütabak,[2] Marina Bazhydai,[3] Ageliki Nicolopoulou,[4] and Paul L. Harris[5]

Abstract

When presented with a claim that contradicts their intuitions, do children seize opportunities to empirically verify such claims or do they simply acquiesce to what they have been told? To answer this question, we conducted a replication of Ronfard et al. (conducted in the People's Republic of China) in two countries with distinct religious and political histories (Study 1: Belarus, $N = 74$; Study 2: Turkey, $N = 79$). Preschool children were presented with five different-sized Russian dolls and asked to indicate the heaviest doll. All children selected the biggest doll. Half of the children then heard a (false) claim (i.e., that the smallest doll was the heaviest), contradicting their initial intuition. The remaining children heard a (true) claim (i.e., that the biggest doll was the heaviest), confirming their initial intuition. Belarusian and Turkish preschoolers typically endorsed the experimenter's claim no matter whether it had contradicted or confirmed their initial intuition. Next, the experimenter left the room, giving children an opportunity to check the experimenter's claim by picking up the relevant dolls. Belarusian and Turkish preschoolers rarely explored the dolls, regardless of the type of testimony they received and continued to endorse the counterintuitive testimony they received. Furthermore, in Study 2, Turkish preschoolers continued to endorse smallest = heaviest even when doing so could have cost them a large reward. In sum, across two different cultural contexts, preschool children endorsed a counterintuitive claim and did not spontaneously seek evidence to test it. These results confirm and extend those of Ronfard et al.

Keywords

Cognitive development, learning, reasoning, counterintuitive, testimony

Humans rely heavily on the accumulated knowledge of their community. Indeed, testimony from other people makes it possible to gather information quickly and to learn about ideas and entities we could not discover on our own. However, other people are not always reliable. They may be ill-informed or ill-intentioned. As a result, individuals have to be epistemically vigilant. They cannot trust everything they are told (Harris, 2012; Sperber et al., 2010). Much recent work has focused on how young children make such evaluations (for reviews see: Harris et al., 2018; Landrum et al., 2015; Mills, 2013; Sobel & Kushnir, 2013). However, this prior work has focused almost exclusively on children's immediate response to testimony: their explicit endorsement or rejection of a given claim (but see Guerrero et al., 2019; Ronfard et al., 2017). This focus makes sense given that much of what children learn from testimony would be difficult or even impossible for them to verify on their own. It is difficult to imagine how a young child (or even an adult!) might seek empirical evidence about the existence of germs, angels, the shape of the earth, the movement of heavenly bodies, or the fall of the Roman Empire. There are, however, many claims that can be empirically tested. For example, if presented with equal-sized cubes and told that some will float and others will sink, children could easily test that claim and learn about the role of density through observation. While we know quite a bit about children's exploration following their *observation* of surprising events (for review see Schulz, 2012), we know very little about their response to surprising *claims*. Do young children seize opportunities to explore surprising claims or do they simply acquiesce to what they have been told?

In an initial study of this question, Ronfard et al. (2018) presented Chinese preschool and elementary school children with five different-sized Russian dolls and asked them to indicate the heaviest doll. As might be expected, almost all children were guided by perceptual clues and pointed to the biggest doll. Half of the children then heard a counterintuitive (and false) claim (i.e., "Actually, the smallest doll is the heaviest"). The remaining children heard a claim confirming their initial intuition (i.e., "Yes, the biggest doll is the heaviest"). Across ages, children typically endorsed the

[1] University of Toronto Mississauga, Canada
[2] Johns Hopkins University, USA
[3] Lancaster University, UK
[4] Lehigh University, USA
[5] Harvard University, USA

Corresponding author:
Samuel Ronfard, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, CCT Building, Room 4059, Mississauga, Ontario L5L 1C6, Canada.
Email: samuel.ronfard@utoronto.ca

experimenter's claim even when it was counterintuitive. However, a developmental difference in children's subsequent *behavior* was clear when the experimenter left the room—preschool children rarely explored the objects, no matter what claim the experimenter had made. By contrast, elementary school children explored the dolls more if the experimenter had made a counterintuitive claim (as opposed to a claim that confirmed their intuition). Thus, at least among Chinese preschoolers, the spontaneous testing of counterintuitive claims develops slowly: when presented with a claim that conflicts with their intuitions, older children evaluate and test it whereas younger children acquiesce, a pattern consistent with developments in children's epistemological thinking, for example, their understanding that factual claims are verifiable, and their ability to design an experiment (Butler et al., 2018; Chen & Klahr, 1999; Koerber & Osterhaus, 2019; Osterhaus et al., 2017).

In the current study, we conducted a replication of Ronfard et al. (2018) to test whether the lack of spontaneous first-hand exploration following a surprising claim observed in Chinese preschoolers is also observed in countries that differ sharply from mainland China: Belarus and Turkey. Belarus and Turkey provide an interesting context to examine whether young children endorse and empirically examine adult claims that are in conflict with their first-hand observation. Belarus' history as a mostly agrarian society and as a member of the Soviet Union means that its citizens endorse conformity, respect for authority, and self-reliance. However, Belarusian society is also connected to Western Europe and its developing national identity prizes individualism (Radzik, 2001). In Turkey, the global spread of urbanization, formal schooling, and socioeconomic development lead Turkish families to adopt a dialectical synthesis of traditional and individualistic values (Kağıtçıbaşı, 1990; 2005). Children's financial contribution to the family is no longer important but parents still preserve their traditional values and psychological interdependence between generations continues. In this regard, showing respect for elders and deference to authority are still highly valued and children are socialized to hold adults' wishes and opinions in high esteem.

Given the tension between independence and respect for authority in both countries, we collected information about parents' endorsement of authoritarian values within each sample to look at whether individual differences in children's endorsement of counterintuitive claims and their exploration of such claims is associated with their parents' socialization goals (Reifen Tagar et al., 2014). Specifically, we tested the hypothesis that preschool children growing up in more authoritarian families may be less likely to seize an opportunity to test an adult's surprising claim than children growing up in less authoritarian families because of the greater respect for authority instilled in them by their parents.

In sum, the current study seeks to replicate and extend the earlier findings with Chinese preschoolers by Ronfard et al. (2018) in two distinct cultures. In doing so, the study would extend the generalizability of the results beyond the Chinese context and offer an opportunity to probe the impact of parental socialization goals on children's exploration—as indexed by the extent to which parents endorse authoritarian values. Finally, by replicating the method and analyses of Ronfard et al. (2018) in Turkey and Belarus, the findings will strengthen the scientific record (Zwaan et al., 2017) and contribute important data from populations usually ignored by developmental researchers (Nielsen et al., 2017), that is, populations that are not from Western, Educated, Industrialized, Rich, and Democratic societies (Henrich et al., 2010)

Table 1. Descriptive statistics by condition (Belarus).

	Counterintuitive testimony	Confirming testimony
Prime	$N = 20$, 11 girls	$N = 17$, 10 girls
	$M_{age} = 4.59$, $SD = .76$	$M_{age} = 4.89$, $SD = .58$
	Age Range = 3.21–5.85	*Age Range* = 3.95–5.77
No prime	$N = 21$, 12 girls	$N = 16$, 11 girls
	$M_{age} = 4.67$, $SD = .85$	$M_{age} = 4.68$, $SD = .72$
	Age Range = 2.98–5.69	*Age Range* = 3.73–5.74

Study 1: Belarus

Data Availability

The data and syntax files for this study are openly available at the Open Science Framework at https://osf.io/wsbuh/?view_only=94e064017cd2414584a60387b2702e98.

Participants

We recruited a total of 86 children from 3 preschools in the city of Minsk, Belarus. Our sample size was designed to match the preschool sample from Ronfard et al. (2018; which was comprised of 81 children) after accounting for attrition. Of these 86 children, 12 children were recruited but not included in our analyses because of equipment failure or experimenter error ($n = 7$) or because children lifted the dolls prior to the experimenter leaving the room ($n = 4$). Thus, our final sample was composed of 74 children (44 female; $M_{age} = 4.70$, $SD = .74$, *Range* = 2.98–5.85; School 1, $n = 25$; School 2, $n = 38$; and School 3, $n = 11$). Preliminary analyses revealed no effect of school on our results.

We randomly assigned children to two conditions: counterintuitive testimony and confirming testimony and then to either a prime to explore or no prime to explore (see Table 1 for descriptive statistics). We obtained a sample that was relatively diverse in family background. Parents reported on the level of education that they and their partner had completed (72 out of 74, or 97% of parents answered this question) and on their income level (71 out of 74, or 96% of parents answered this question). Of the parents who responded, 25% reported that neither parent had earned a college degree and 75% reported that at least one parent had earned a college degree. Parents reported having: a higher-income level (1%), a middle-income level (86%), a lower-income level (1%), or did not want to report their income level (12%). The surveys were mostly completed by children's mothers (86% mothers and 14% fathers).

This study was approved by the Ethics Committees of Harvard University (IRB#1242). Parents of participants gave informed consent in writing before children participated in the study and children gave verbal assent.

Materials

We used five, different-sized, Russian nesting dolls; each doll was attached to a square base for stability. Size and weight were correlated—the smallest doll was the lightest, and the biggest doll was the heaviest. The dolls and their bases were painted white. They were arranged on a tray placed on the table so that the biggest doll was on the child's left and the smallest doll was on the child's right (Figure 1). The experimenter and the child sat next to each other at

Figure 1. Stimuli used in the experiment.

the table. The dolls were approximately 18″ from the table edge nearest to the child. The experimental session was discreetly recorded using a laptop camera with a darkened screen.

Procedure

The six-phase procedure for this study was identical to that used by Ronfard et al. (2018). Children were individually tested in a separate room at their school by a female Belarusian experimenter fluent in Russian—the native language of the children and the language used in the school.

Initial Judgment

The experimenter asked children to point to the heaviest doll: "Which doll do you think is the heaviest?" Children were then asked an open-ended question inviting them to explain their judgment: "You think this one is the heaviest—why do you think it is the heaviest?"

Testimony

Children were randomly assigned to receive either counterintuitive testimony (i.e., smallest = heaviest) or testimony that confirmed their intuition (i.e., biggest = heaviest). In the counterintuitive testimony condition, the experimenter told children:

> Actually, that one is not the heaviest; this one here (pointing to the smallest one on the right) is the heaviest. It is heavier than all of the other ones. It's heavier than this one, this one, this one, this one (starting with the biggest one and moving to the second smallest one).

Note that this statement was false because, fully consistent with the appearance of the dolls, the smallest doll was the lightest and the biggest doll was the heaviest. In the confirming testimony condition, the experimenter told children:

> Yes, that one is the heaviest, and this one here (pointing to the smallest one on the right) is the lightest. This one (pointing to the biggest one) is heavier than all of the other ones. It is heavier than this one, this one, this one, and this one (starting with the second largest one and moving to the smallest one).

Post-Testimony Judgment

Children were again asked to identify the heaviest doll and to provide an explanation for their judgment using the same wording as for the initial judgment: "Which doll do you think is the heaviest?" Children were also asked to recall which doll the experimenter had identified as the heaviest: "Can you point to the one I said was the heaviest?"

Opportunity to Explore the Dolls

For children in the no prime condition, the experimenter then told children that she was going next door to use her phone for a moment but that she would come right back. For children assigned to the prime condition, she added, just before leaving: "I'll move the dolls a bit closer to you" and pushed the tray so that the dolls were about 6 inches from the child. She then walked out of the room and returned after 60 s.

Opportunity to Report Exploration to the Experimenter

Once the experimenter returned, she said, "Let's see—we were talking about the dolls," and paused for 10 s to offer children an opportunity to initiate a conversation with her following their opportunity to explore the dolls. If children did not spontaneously comment, she prompted children: "Okay, we're almost done. Is there anything you want to tell me?"

Final Judgment

Children were again asked to identify the heaviest doll and to provide an explanation for their judgment using the same wording as the initial judgment.

Coding

To investigate children's exploration, we coded from the video how many times each child picked up each of the five dolls. Two research assistants, blind to the hypotheses of the study, coded 100% of the videos for children's exploration of the doll. Both coders were also blind to children's age, condition, and judgments about the dolls. Inter-rater agreement, as measured by Cohen's Kappa (κ) for each doll, was excellent: smallest doll, $\kappa = .90$;

Table 2. Percentage of Belarusian children in each condition who claimed that the biggest doll was the heaviest at three successive time-points.

Timing of judgment	Confirming ($N = 33$; %)	Counter ($N = 41$; %)
Initial	100	100
Post-testimony	100	13
Post-opportunity to spontaneously explore	97	27

second smallest doll, $\kappa = .93$; middle doll, $\kappa = .83$; second biggest doll, $\kappa = .89$; and biggest doll, $\kappa = .94$.

Measure of Parent's Endorsement of Authoritarian Values

Parents were asked to indicate in four forced-choice items which of two child-rearing values (authoritarian vs. non-authoritarian) they found more important (Feldman & Stenner, 1997; Stenner, 2005). The value pairs were "independence" versus "respect for elders," "obedience" versus "self-reliance," "curiosity" versus "good manners," and "being considerate" versus "[being] well-behaved." After scoring (1 = *authoritarian response*, 0 = *non-authoritarian response*), responses were averaged to run from 0 to 1, with higher scores indicating a stronger authoritarian predisposition. When parents did not complete all 4 items, a pro-rated score based on the number of items completed was created. Responses were obtained from 67 out of 74 parents (91%), $M = .29$, $SD = .34$, and *Range* = 0–1. The reliability of the scale, computed based on the 44 children for whom parents had answered all four questions, was Cronbach's alpha (α) = .37.

Results

We analyze children's: (i) initial and post-testimony weight judgments, (ii) exploration of the dolls, (iii) post-exploration weight judgments, and (iv) post-exploration weight judgments as a function of children's exploration. Neither parental authoritarianism nor children's age-predicted variability in children's judgments and exploration; details of these analyses can be found in Online Supplementary Materials.

Children's Initial and Post-Testimony Weight Judgments

As Table 2 shows, all children initially stated, as expected, that the biggest doll was the heaviest. Analysis of children's explanations confirmed that children associated greater size with greater weight: 77% of children mentioned size as a justification of their selection whereas the remaining children either did not provide an explanation (19%) or provided an explanation unrelated to the size of the dolls (4%).

Among those children who heard testimony confirming their initial judgment, 32 of 33 children continued to make the same judgment post-testimony and one child failed to provide a response, McNemar Test (binomial test for this and all subsequent tests) = 1.0. In contrast, among children who heard testimony conflicting with their initial judgment only 5 of 41 continued to make the same judgment post-testimony and two children failed to provide a response. Thus, few children persisted with their initial judgment in the counterintuitive testimony condition, McNemar Test, $p < .001$. They justified their response by simply repeating what they had been told (44%, "The smallest is the heaviest"), failed to provide a justification (41%), or provided other unrelated explanations (15%, e.g., "It's the middle one"). Comparing across conditions, children who received counterintuitive testimony endorsed the biggest doll as the heaviest significantly less often than children who received confirming testimony, $\chi^2(1, N = 71) = 53.53$, $p < .001$, Cramér's $V = .86$. Thus, the type of testimony children received markedly affected their identification of the heaviest doll.

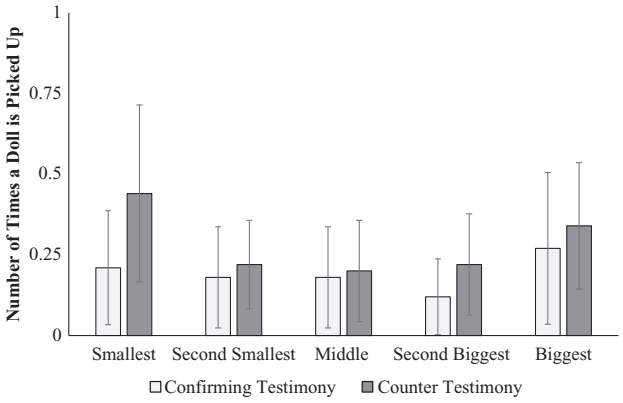

Figure 2. Number of times each doll was picked up by Belarusian preschoolers when the experimenter left the room by condition. Error bars represent 95% confidence intervals. Confirming Testimony Condition ($N = 33$) and Counter Testimony Condition ($N = 41$).

Children's Exploration of the Dolls

In Figure 2, we display the mean number of times children picked up each doll by condition. We conducted a mixed analysis of variance (ANOVA) with the between-subject factors of Testimony Type (2: Counterintuitive vs. Confirming), and Priming (2: Prime vs. no Prime), and the within-subject factor of Doll (5: one [i.e., smallest], two, three, four, and five [i.e., biggest]) on the number of times children picked up a doll. This analysis revealed an effect of Doll, $F(4, 280) = 3.59$, $p = .007$, and $\eta_p^2 = .05$. Children tended to pick up the smallest and the biggest dolls more frequently than the dolls of intermediate sizes. Critically, however, there was no effect of Testimony or of receiving a Prime or significant interactions. By implication, counterintuitive testimony did not provoke Belarusian preschool children to test the experimenter's claim. This lack of selective exploration does not seem to reflect shyness on the part of children. Children rarely explored the dolls even when they were primed to explore and had received testimony that the smallest = heaviest. Indeed, the percentage of children who picked up at least one doll was low across all conditions: Counter Testimony No Prompt = 33% (7 out of 21), Counter Testimony Prompt = 25% (5 out of 20), Confirming Testimony No Prompt = 13% (2 out of 16), and Confirming Testimony Prompt = 29% (5 out of 17).

Post-Exploration Weight Judgments

Following the return of the experimenter, few children in the counterintuitive and confirming testimony conditions commented on the weight of the dolls or on the fact that they had picked up the dolls (see Online Supplementary Materials). When the experimenter

asked children to identify the heaviest doll a final time, a clear condition emerged (Table 2). Children who had received testimony confirming their intuitions that the biggest doll was the heaviest continued to claim that the biggest doll was the heaviest. There was a modest change in children's endorsement of the smallest = heaviest in the counterintuitive testimony condition, but this fell short of significance, McNemar Test, $p = .07$. Moreover, at the end of the experiment, the type of testimony children had received continued to markedly affect their judgments of the doll's weights, $\chi^2(1, N = 71) = 36.95$, $p < .001$, and Cramér's $V = .71$.

Post-Exploration Weight Judgments as a Function of Children's Exploration

To provide a more targeted assessment of whether children's exploration impacted their subsequent weight judgments, we analyzed those judgments as a function of whether children had explored the dolls in the experimenter's absence, restricting our analysis to children who had received counterintuitive testimony. Following the coding scheme of Ronfard et al. (2018), we operationalized exploration as children's decision to lift the biggest *and* the smallest doll (picked up one after another, i.e., not simultaneously). As compared to children who did not explore in this fashion, a greater percentage of children who did explore stated that the biggest doll was the heaviest on their final judgments (23% vs. 40%) but this difference was not statistically significant, $\chi^2(1, N = 41) = 1.17$, $p = .28$, and Cramér's $V = .17$.

Study 1: Summary

Overall, these data indicate that preschool children in Belarus, like their peers in China (Ronfard et al., 2018), used perceptual clues to infer that the biggest doll was the heaviest doll. Nevertheless, when provided with a claim running counter to that intuition, most children accepted that claim. In addition, they were no more likely to seek out empirical evidence following that counterintuitive claim than following a claim confirming their intuitions. When surveying the overall pattern of children's responses across the successive steps of the experimental procedure in the counterintuitive condition only, four groups emerged. The majority of children (54%, $n = 22$) endorsed the experimenter's claim, did not test it, and continued to endorse it at the posttest. Another 15% ($n = 6$) of children did not test the experimenter's claim but vacillated in their endorsement of it, for example endorsing it immediately after hearing it but not when the experimenter returned to the room or vice versa. A very small group of children (7%, $n = 3$) were skeptical of the claim initially (did not endorse it), but did not test it, and did not endorse it on the experimenter's return. The last group of children (25%, $n = 10$) did check the experimenter's claim by investigating the dolls but fewer than half of those children (i.e., only 10%, $n = 4$, of the children in the counterintuitive condition) went on to state that the biggest doll was the heaviest when the experimenter returned to the room.

In Study 2, we tested Turkish preschoolers on the same task with one noteworthy addition. To assess whether Turkish preschoolers truly believed the experimenter's claim that the smallest doll was the heaviest, we presented children with a large candy bar on one side of a balance scale. We told children that only a very heavy object could lift the candy bar and told them that if they selected a doll heavy enough to lift the candy bar they would be able to keep it. We reasoned that if children were simply pretending to endorse the claim that the smallest doll was the heaviest then they would be significantly more likely to select the biggest doll on this task relative to the prior task when the experimenter asked them which doll was the heaviest.

Table 3. Descriptive statistics by condition (Study 2—Turkey).

Counterintuitive testimony	Confirming testimony
$N = 39$, 21 girls	$N = 40$, 21 girls
$M_{age} = 4.41$, $SD = .65$	$M_{age} = 4.43$, $SD = .75$
Age Range = 3.28–5.95	*Age Range* = 3.14–6.00

Study 2: Turkey

Data Availability

The data and syntax files for this study are openly available at the Open Science Framework at https://osf.io/wsbuh/?view_only=94e064017cd2414584a60387b2702e98.

Participants

We recruited a total of 89 children from 6 preschools in the city of Kayseri, Turkey. Our sample size was designed to match the preschool sample from Ronfard et al. (2018; which was comprised of 81 children) after accounting for attrition. Of these 89 children, 10 children were not included in our analyses: 2 because of experimenter error, 2 because they did not initially identify the biggest doll as the heaviest, and 6 because of equipment failure. As a result, our final sample was composed of 79 children (42 girls; $M_{age} = 4.42$, $SD = .70$, *Range* = 3.14–6.00; School 1, $n = 25$; School 2, $n = 18$; School 3, $n = 15$; School 4, $n = 13$; School 5, $n = 5$; and School 6, $n = 3$). Preliminary analyses revealed no effect of school on our results. We randomly assigned children to two conditions: counterintuitive testimony and confirming testimony (see Table 3 for descriptive statistics). Given the absence of any effect of prompt in Study 1, we did not assign children to different prompt conditions in Study 2.

We obtained a sample that was relatively highly educated in family background. All parents reported on the level of education they and their partner had completed and on their monthly expenses (as a proxy for family income). We made the decision to ask about monthly expenses rather than family income because Turkish families were hesitant to give information about their income level. Of the parents who responded: 10% reported that neither parent had completed college, 84% reported at least one parent had a college degree, and 6% reported that at least one parent had a graduate degree (Master or Doctorate). Parents were also asked to report on their monthly expenses (minimum wage at the time of testing was 2000 Turkish Liras): 5,000 or more (48%), 3,000–5,000 (32%), 1,200–3,000 (19%), and less than 650 (1%).

This study was approved by the Ethics Committees of XXXX University (IRB#1242). Parents of participants gave informed consent in writing before children participated in the study and children gave verbal assent.

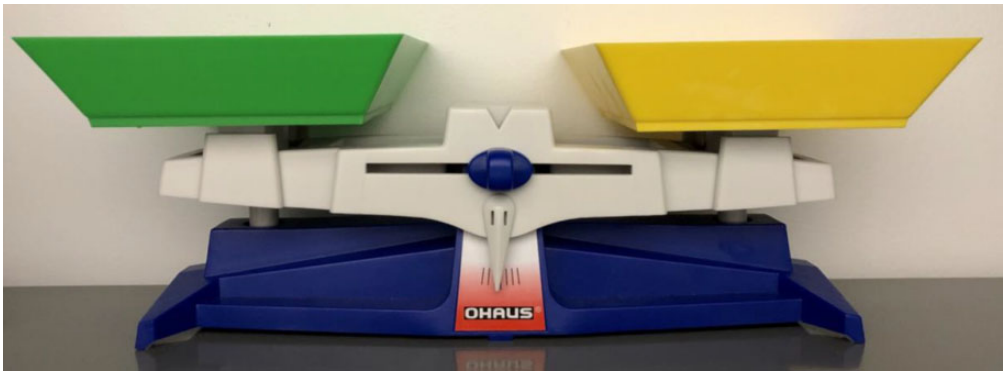

Figure 3. Balance scale similar to the one used in the final phase of Study 2.

Procedure

The procedure for this study was identical to Ronfard et al. (2018) and to Study 1. Children were individually tested in a separate room at their school by a female Turkish experimenter fluent in Turkish. As before, the experimental procedure consisted of six phases. However, at the end of these six phases, an additional phase was added—a balance scale task (see Figure 3). Children were presented with a balance scale and were told:

> I have a fun game for you to play. This is a scale. On this side, there is a box with candy. To get the candy you have to put the heaviest doll on this side so that it lifts the box of candy that is on this side. If it is not heavy, it will not lift the box of candy and you won't get any candy. You will get the candy, if you put something very heavy here that lifts the candy on this side.

Children were then asked to confirm that they understood the point of the game: "OK, so what do you need to do to get the candy in this box?" If they correctly answered, they were told: "Right, you need to put something very heavy on this side." If they responded incorrectly, the game was explained to them again and the confirmation question was asked a second time. Children were then asked to select a doll: "OK, go ahead. Put the doll you think is the heaviest on this side." The doll they selected was then placed on the balance scale so children could observe whether it was heavy enough. Note that only the biggest doll was heavy enough to tip the scale. If the child did not select the biggest doll, they were given another turn: "OK, it looks like this doll was not heavy enough. Go ahead and try another doll."

Coding

The same coding scheme as in Study 1 was used to code children's exploration of the dolls. Two research assistants, blind to the hypotheses of the study, coded 100% of the videos for children's exploration of the doll. Both coders were blind to children's age, condition, and judgments about the dolls. Inter-rater agreement as measured by Cohen's κ was excellent for each doll: smallest doll, $\kappa = .84$; second smallest doll, $\kappa = 1.00$; middle doll, $\kappa = 1.00$; second biggest doll, $\kappa = .89$; and biggest doll, $\kappa = .95$.

Measure of Parent's Endorsement of Authoritarian Values

The same measure of parental authoritarianism described in Study 1 was used for Study 2. Responses were obtained from 74 out of 79 parents (94%), $M = .38$, $SD = .28$, and $Range = 0–1$. The reliability of the scale computed based on the 74 children for whom parents had answered all four questions was Cronbach's $\alpha = .39$.

Table 4. Percentage of Turkish children in each condition who claimed that the biggest doll was the heaviest at six successive time-points.

Timing of judgment	*Confirming*	*Counter*
Initial	100% ($n = 40$)	100% ($n = 39$)
Post-testimony	93% ($n = 40$)	10% ($n = 39$)
Post-opportunity to spontaneously explore	98% ($n = 40$)	28% ($n = 39$)
Balance task—first attempt	95% ($n = 40$)	33% ($n = 39$)
Balance task—second attempt ($n = 27$)	100% ($n = 2$)	64% ($n = 25$)
After both attempts ($n = 79$)	100% ($n = 40$)	77% ($n = 39$)

Note. One child in the counterintuitive testimony condition who did not select the biggest doll on the first attempt did not complete the second attempt. Thus, 25 out of the 26 children the counterintuitive testimony condition who did not select the biggest doll in the first attempt completed the second attempt.

Results

We analyze children's: (i) initial and post-testimony weight judgments, (ii) exploration of the dolls, (iii) post-exploration weight judgments, (iv) post-exploration weight judgments as a function of children's exploration, and (v) children's doll selection on the balance scale task. As in Study 1, our analyses indicated that parental authoritarianism and children's age did not predict variability in children's judgments and exploration. Details of these analyses can be found in Online Supplementary Materials.

Children's Initial and Post-Testimony Weight Judgments

As Table 4 shows, all children initially stated that the biggest doll was the heaviest. Analysis of children's explanations confirmed that children associated greater size with greater weight: 59% of children mentioned size as a justification of their selection of the biggest doll as the heaviest while the remaining children either did not provide an explanation (9%) or provided an explanation unrelated to the size of the dolls (32%).

When asked to make the post-testimony judgment, all but three children (90%) who heard testimony that confirmed their initial judgment made the same judgment, McNemar Test, $p = = .25$. In contrast, very few of the children (4 out of 39) who heard

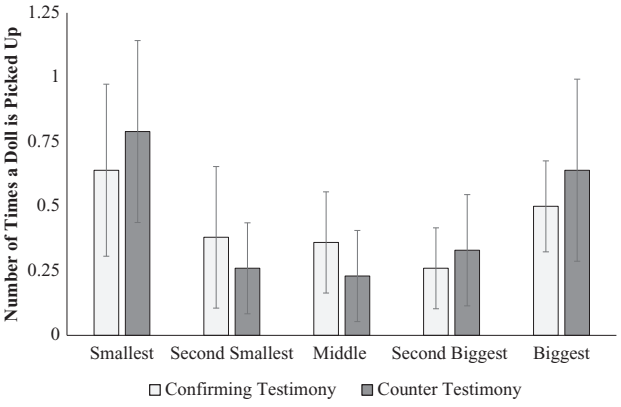

Figure 4. Number of times each doll was picked up by Turkish preschoolers when the experimenter left the room by condition. Error bars represent 95% confidence intervals. Confirming Testimony Condition ($N = 40$) and Counter Testimony Condition ($N = 39$).

testimony that conflicted with their initial judgment continued to make the same judgment, McNemar Test, $p < .001$. The majority now stated that the smallest was the heaviest. Just over half of these children justified their response by simply repeating what they had been told (51.5%, "The smallest is the heaviest"). The remaining children either did not provide a justification (11.5%) or provided a justification unrelated to the testimony they received (27%, e.g., "It came to my mind"). Children who received counterintuitive testimony endorsed the biggest doll as the heaviest significantly less often than children who received confirming testimony, $\chi^2(1, N = 79) = 53.50$, $p < .001$, and Cramér's $V = .82$. Thus, the type of testimony children received markedly affected their post-testimony judgments of the dolls' weights.

Children's Exploration of the Dolls

In Figure 4, we display the mean number of times children picked up each doll by condition. We conducted a mixed ANOVA with the between-subject factors of Testimony Type (2: Counterintuitive vs. Confirming) and the within-subject factor of Doll (5: one [i.e., smallest], two, three, four, and five [i.e., biggest]) on the number of times children picked up a doll. This analysis revealed an effect of Doll, $F(4, 308) = 10.28$, $p < .001$, and $\eta_p^2 = .12$. Children lifted the smallest and the biggest dolls at equal rates but lifted both more frequently than the dolls of intermediate sizes, all $p < .038$. Critically, however, there was no effect of Testimony and no interaction of Testimony × Doll. Counterintuitive testimony did not provoke Turkish preschool children to test the experimenter's claim by lifting the dolls more often. The percentage of children who picked up at least one doll was low across both conditions: Counter Testimony = 49% (19 out of 39) and Confirming Testimony = 50% (20 out of 40).

Post-Exploration Weight Judgments

Following the return of the experimenter, few children in the counterintuitive and confirming testimony conditions commented on the weight of the dolls or on the fact that they had picked up the dolls (see Online Supplementary Materials). When the experimenter asked children about the weight of the doll a final time, a clear effect of condition emerged. Children who had received testimony confirming their intuition that the biggest doll was the heaviest continued to endorse the biggest doll as the heaviest in the post-exploration phase. In contrast, although the majority of children who had received testimony that the smallest = heaviest continued to endorse that testimony in the post-exploration phase, more children claimed that biggest = heaviest at the end of the experiment than did so immediately after having received the counterintuitive testimony, McNemar Test, $p = .039$. Nevertheless, in the post-exploration phase, the type of testimony children had received continued to markedly affect their judgments of the doll's weights, $\chi^2(1, N = 79) = 40.81$, $p < .001$, and Cramér's $V = .72$.

Post-Exploration Weight Judgments as a Function of Children's Exploration

To provide a more targeted assessment of whether children's exploration impacted their subsequent weight judgments, we analyzed those judgments as a function of whether they had explored the dolls in the experimenter's absence, restricting our analysis to children who had received counterintuitive testimony. We again operationalized exploration as children's decision to pick up the biggest *and* the smallest doll at some point during the experimenters' absence. As compared to children who did not explore, a greater percentage of children who did explore stated that the biggest doll was the heaviest on their final judgments (18% vs. 55%), $\chi^2(1, N = 39) = 5.25$, $p = .022$, and Cramér's $V = .37$.

Children's Doll Selection on the Balance Scale Task

At the end of the interview, children were told that they would earn a bag of candy if they selected a doll heavy enough to lift a bag of candy on the balance scale. Choosing the biggest doll was coded as 1 while choosing any other doll was coded as 0. On their first attempt, children's selection of which doll to put on the scale differed significantly by condition. Almost all children in the confirming testimony condition selected the biggest doll whereas only a third of children who had received counterintuitive testimony did so, 95% versus 33%, $\chi^2(1, N = 79) = 32.82$, $p < .001$, and Cramér's $V = .64$. Of the 26 children in the counterintuitive condition (67% of children in that condition) who did not select the biggest doll, 23 selected the smallest doll—the doll the experimenter had claimed was the heaviest. Children who did not initially select the biggest doll and thus were not able to earn the bag of candy were given a second chance. Importantly, these children now had evidence that the smaller doll they had selected was not heavy enough and could use this evidence to change their selection for their second attempt. By the end of this second round, 100% of the children in the confirming testimony condition had selected the biggest doll. In contrast, by the end of the second round, 23% of children in the counterintuitive testimony condition still had not selected the biggest doll (Table 4).

Study 2: Summary

Preschool children in Turkey, like their peers in Belarus (Study 1) and China (Ronfard et al., 2018), are no more likely to seek out empirical evidence following a counterintuitive claim than following a claim that confirms their intuitions. Indeed, when looking at the pattern of children's responses in the counterintuitive condition only ($N = 39$), Turkish children look very similar to Belarusian

children. The majority (56%, $n = 22$) endorsed the experimenter's testimony, did not test it, and continued to endorse it at the posttest; 10% ($n = 4$) of children did not test the experimenter's claim but vacillated in their endorsement of it. A very small group of children (5%, $n = 2$) were skeptical of the claim initially (did not endorse it), did not test it, and did not endorse it on the experimenter's return. The last group of children (28%, $n = 11$) did assess the experimenter's claim by exploring the dolls with about half of those children stating that biggest = heaviest (i.e., only 15%, $n = 6$ of the children in the counterintuitive condition) when the experimenter returned to the room.

Discussion

Do preschool children seize opportunities to empirically examine surprising claims—claims that run counter to their intuitions? To answer this question, we conducted a replication of Ronfard et al. (2018; conducted in the People's Republic of China) in Belarus and Turkey. Replicating prior work, we found that most children endorsed the counterintuitive claim that the smallest doll was the heaviest (Belarus = 87% and Turkey = 90%) and did not spontaneously investigate it by picking up the dolls in the experimenter's absence (Belarus = 75% and Turkey = 72%). Extending prior work, we found that children continued to endorse the claim that the biggest doll was the heaviest even at a potential cost to themselves: When asked to select a doll that would be heavy enough to lift a large bag of candy, the majority of Turkish children selected the smallest doll. However, on subsequent attempts, they often switched to selecting the biggest doll. By implication, preschool children can update their belief in the experimenter's claim when faced with empirical evidence contracting that claim. However, they do not appear to spontaneously seek out such evidence. Also extending prior work, we found that whether or not children tested the experimenter's claim was unrelated to their parents' endorsement of authoritarian beliefs (Experiments 1 and 2). In sum, across diverse countries, preschool children generally accept, and are willing to act on the basis of, an adult's surprising claim without testing the truth of that claim. Why?

One reason for preschool children's apparent credulity may be the nature of claim. Weight is an invisible property. Children do not have direct perceptual access to it. Moreover, although preschool children expect bigger objects to be heavier, they also know that some small objects can be heavy and that some big objects can be light. Our claim was moderately counterintuitive and this likely contributed to children's endorsement of it (see Lane, 2018 for a review). Yet, by itself, the nature of the claim cannot explain why, unlike preschool children, elementary school children engage in empirical testing of that claim (Ronfard et al., 2018; under review).

One explanation for the age-related difference in children's spontaneous exploration of the informant's claim is that preschool and elementary school children differ in their perception of the informant's reliability. Unlike preschool children, elementary school children likely understand that speakers sometimes make false claims—because they are being ironic, cynical, or mendacious (Filippova & Astington, 2008; Mills & Kiel, 2005; Peterson et al., 2012; Talwar & Lee, 2008) or alternatively because their representation of the world is mistaken (Astington et al., 2002). As a result, elementary school children may take a more cautious approach to surprising claims than preschool children. If age-related differences in children's empirical testing are a result of differences in children's perceptions of informant reliability, then preschool children's empirical testing should increase when they are faced with an informant who is clearly unsure about his or her claim, "I know" versus "I think but I'm unsure."

An alternative or additional reason for the lack of empirical testing of counterintuitive claims by preschoolers is that testing a counterintuitive claim requires children to reason through the empirical implications of the claim and to realize how those implications could be checked and potentially falsified. Effectively, our task may tap into children's developing ability to design unconfounded experiments—a skill that emerges around the elementary school years (Chen & Klahr, 1999) and that is related to developments in children's epistemological thinking (Koerber & Osterhaus, 2019; Osterhaus, et al., 2017). Indeed, it may be that the spontaneous decision to test a counterintuitive claim necessitates an understanding that factual claims are verifiable (Butler, et al., 2018).

So far, we have only considered two cognitive explanations for preschool children's failure to test the experimenter's claim. However, it is possible that preschoolers simply did not feel that it was appropriate to carry out such a test. After all, children are often told not to touch things that do not belong to them. However, this explanation faces two challenges. First, receiving a prime to explore did not increase exploration by preschool children who received counterintuitive testimony in Belarus (Experiment 1), Hong Kong, the U.S. (Ronfard et al., under review), and China (Ronfard et al., 2018). However, it did impact the empirical testing of elementary school children in the U.S. and Hong Kong (Ronfard et al., under review). By implication, only elementary school children's empirical testing seems to be inhibited by concerns about permissibility—their exploration increases when they are prompted to explore, but this is not the case for preschool children. Second, parental authoritarianism was not associated with preschool children's exploration of the experimenter's claim in Belarus, Turkey, the U.S., and Hong Kong (Ronfard, under review) but it was associated with the empirical testing of elementary school children in the U.S. and Hong Kong (Ronfard et al., under review). Thus, while the low reliability of the parental authoritarianism scale is a concern and warrants additional research, it does predict empirical testing by older children. This suggests that sociocultural factors and children's belief that they are allowed to test the claim do influence children's empirical testing but apparently only after they have developed the cognitive resources necessary to doubt such claims, as proposed in the preceding paragraph. Thus, researchers interested in how parents and schools influence children's responses to counterintuitive claims should consider investigating such effects with older children. For example, researchers might investigate whether pedagogy that emphasizes children's independent learning (as in Montessori and Reggio Emilia schools) rather than their trust in a teachers' testimony is associated with increased empirical testing of informants' claims.

In sum, we asked: do preschool children seize opportunities to empirically check counterintuitive claims or do they simply acquiesce to what they have been told? Across diverse cultures, they acquiesce. Future work will need to investigate the breadth of this effect as well as its underlying cause. Paradoxically, it may turn out that young children's failure to seek empirical evidence following counterintuitive claims is adaptive rather than maladaptive. It allows them to quickly accept opaque, hard-to-discover, and counterintuitive information from teachers and caregivers. Children who insisted on empirically checking every counterintuitive claim

before accepting its veracity might display the type of cognitive autonomy eulogized by Rousseau in his classic account of Emile's education, but such children would be stubborn misfits from a pedagogical standpoint.

Acknowledgments

We appreciate the school staff for their support, and we thank the parents and children for their participation. Data collection in Belarus was funded by a Mind, Brain, and Behavior Graduate Student Award from Harvard University to Samuel Ronfard.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Samuel Ronfard https://orcid.org/0000-0001-5152-0514

Supplementary Materials

Supplemental material for this article is available online.

References

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, *20*, 131–144. https://doi.org/10.1016/S0732-118X(02)00005-3

Butler, L. P., Schmidt, M. F., Tavassolie, N. S., & Gibbs, H. M. (2018). Children's evaluation of verified and unverified claims. *Journal of Experimental Child Psychology*, *176*, 73–83. https://doi.org/10.1016/j.jecp.2018.07.007

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*, 1098–1120. https://doi.org/10.1111/1467-8624.00081

Feldman, S., & Stenner, K. (1997). Perceived threat and authoritarianism. *Political Psychology*, *4*, 741–770.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*, 126–138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Guerrero, S., Sebastián-Enesco, C., Pérez, N., & Enesco, I. (2019). Myths in science: Children trust but do not retain their teacher's information. *Journal of Applied Developmental Psychology*, *62*, 116–121. https://doi.org/10.1016/j.appdev.2019.02.007

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Belknap Press/Harvard University Press.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Kağıtçıbaşı, Ç. (1990). Family and socialization in cross-cultural perspective: A model of change. In J. Berman (Ed.), *Nebraska symposium on motivation, 1989* (pp. 135–200). University of Nebraska Press.

Kağıtçıbaşı, Ç. (2005). Autonomy and relatedness in cultural context: Implications for self and family. *Journal of Cross-Cultural Psychology*, *36*, 403–422. https://doi.org/10.1177/0022022105275959

Koerber, S., & Osterhaus, C. (2019). Individual differences in early scientific thinking: Assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development*, 510–533. https://doi.org/10.1080/15248372.2019.1620232

Landrum, A. R., Eaves, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences*, *19*, 109–111. https://doi.org/10.1016/j.tics.2014.12.007

Lane, J. D. (2018). Children's belief in counterintuitive and counterperceptual messages. *Child Development Perspectives*, *12*, 247–252. https://doi.org/10.1111/cdep.12294

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*, 404. https://doi.org/10.1037/a0029500

Mills, C. M., & Keil, F. C. (2005). The development of cynicism. *Psychological Science*, *16*, 385–390. https://doi.org/10.1111/j.0956-7976.2005.01545.x

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

Osterhaus, C., Koerber, S., & Sodian, B. (2017). Scientific thinking in elementary school: Children's social cognition and their epistemological understanding promote experimentation skills. *Developmental Psychology*, *53*, 450–462. https://doi.org/10.1037/dev0000260

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*, 469–485. https://doi.org/10.1111/j.1467-8624.2011.01728.x

Radzik, R. (2001). Belarus between the east and the west. *International Journal of Sociology*, *31*, 11–45.

Reifen Tagar, M., Federico, C. M., Lyons, K. E., Ludeke, S., & Koenig, M. A. (2014). Heralding the authoritarian? Orientation toward authority in early childhood. *Psychological Science*, *25*, 883–892. https://doi.org/10.1177/0956797613516470

Ronfard, S., Chen, E. E., & Harris, P. L. (2018). The emergence of the empirical stance: Children's testing of counter-intuitive claims. *Developmental Psychology*, *54*, 482–493. https://doi.org/10.1037/dev0000455

Ronfard, S., Chen, E. E., & Harris, P. L. (under review). Testing what you're told: Older American and Chinese children empirically verify a counterintuitive claim.

Ronfard, S., Lane, J.D., Wang, M., & Harris, P.L. (2017). The impact of counterperceptual testimony on children's categorization after a delay. *Journal of Experimental Child Psychology*, *163*, 151–158. https://doi.org/10.1016/j.jecp.2017.06.006

Schulz, L.E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Science*, *16*, 382–389.

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*, 779–797. https://doi.org/10.1037/a0034191

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*, 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stenner, K. (2005). *The authoritarian dynamic*. Cambridge University Press.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, *79*, 866–881. https://doi.org/10.1111/j.1467-8624.2008.01164.x

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/s0140525x17001972