Violations of Core Object Principles Change Adults' Behaviors in Maze Games

Rongzhi Liu (rongzhi @uchicago.edu)
Department of Psychology
University of Chicago
Chicago, IL 60637 USA

Fei Xu (fei_xu@berkeley.edu)
Department of Psychology
University of California, Berkeley
Berkeley, CA 94720 USA

Abstract

A set of fundamental principles governs our reasoning about objects since infancy: solidity, continuity, and contact. Past studies have shown that adults can revise these principles given a small amount of counterevidence. However, how far would they generalize their revised beliefs? In the present experiments, we demonstrate that given a diverse set of counterevidence, adults changed their behaviors in subsequent maze games. These results demonstrate that adults can generalize their revised beliefs about the core object principles to a completely different virtual environment.

Keywords: belief revision; core knowledge; intuitive physics

Introduction

Humans are remarkable learners. We have invented complex technologies such as smartphones, computers, and artificial intelligence. More impressively, we can overturn our previous beliefs and develop completely new beliefs. Einstein's theory of general relativity replaced Newtonian mechanics, changing what people believed about how the universe worked for hundreds of years. Indeed, many cognitive scientists and developmental psychologists have argued that one of the hallmarks of human intelligence is that we form beliefs and revise them given new evidence (Chater & Oaksford, 2008; Gopnik & Wellman, 2012; Piaget, 1954; Tenenbaum et al., 2011; Ullman & Tenenbaum, 2020; Xu, 2019). Past research has shown that adults rationally revise their beliefs given new evidence in many domains, including objects (Allen et al., 2020; Hamrick et al., 2016), agents (Baker et al., 2017, Jara-Ettinger et al., 2020), and causal reasoning (Lucas & Griffiths, 2010; Griffiths et al., 2011). However, are there limits to humans' ability to rationally revise our beliefs? Are there beliefs that are so entrenched that cannot be revised?

Some have argued that the most fundamental domains of knowledge are the core knowledge systems (Spelke, 1988, 2000, 2022; Spelke & Kinzler, 2007). The core knowledge systems are a small number of systems of domain-specific knowledge, each accompanied by a set of principles. Spelke (2022) has proposed 6 core knowledge systems: places, objects, number, forms, agents, and social beings. These systems emerge early in development (Baillargeon, 2008; Spelke, 2022), are universal in humans across cultures (Barrett et al, 2013; Gordon, 2004), and are shared with some nonhuman animals (Hare et al., 2001; Regolin & Vallartigara, 1995). Thus, beliefs in the core knowledge systems are the most entrenched beliefs that humans hold since infancy.

One of the most richly studied core knowledge systems is the system of objects, which guides how we represent and reason about objects and their motions. The principles in this system include solidity – objects cannot occupy the same space at the same time (Spelke et al., 1992), continuity – objects exist and move continuously in time and space (Anguiar & Baillargeon, 1999), cohesion – objects move as connected and bounded wholes (Anguiar & Baillargeon, 1999), and contact – objects do not interact at a distance (Leslie & Keeble, 1987). These principles emerge by 2.5 to 6 months of age, they support further learning in the physical domain (Baillargeon, 2008), and they persist into adulthood (Scholl & Pylyshyn, 1999; Scholl et al., 2001; vanMarle & Scholl, 2003).

Are these earliest-emerging core principles about objects also subject to revision? Recent studies have shown that adults and preschoolers can revise their beliefs about the solidity, continuity, and contact principles in a virtual environment, given just a small amount of counterevidence (e.g., balls going through a solid wall) (Liu & Xu, 2021, 2022, under review). However, an open question is how far learners will generalize their revised beliefs about the core object principles. Would they generalize their revised beliefs to completely different events in completely different environments? Furthermore, when adults were asked to explain the counterevidence they observed, some adults readily accepted the violations, whereas others came up with alternative interpretations to explain away the violations (e.g., there was a hole in the wall that the balls went through). Yet a third group of adults mentioned that they were simply tracking the statistical patterns of the counterevidence (Liu & Xu, 2022, under review). Although the latter two groups of participants still predicted outcomes inconsistent with the principles for new events, it is unclear whether they have genuinely accepted the counterevidence and revised their beliefs about the core object principles.

In the present study, we investigate these open questions by giving participants a completely different task after they observe evidence about the core object principles. Specifically, participants observed 6 events that supported or violated each principle, or they did not receive any new evidence about the principle. Then participants played a maze game, in which they chose between actions consistent with the principle or actions violating the principle. This task allows us to examine whether participants would generalize their revised beliefs about the principles to a completely different environment. In addition, since the choices in the maze games were completely different from the events they observed in the counterevidence, this design reduces the likelihood that participants would choose actions inconsistent with the principles based on alternative interpretations of the counterevidence (since the alternative interpretations would not apply to the maze games) or by simply tracking statistical regularities in the counterevidence. We tested three core object principles: Solidity, Continuity, and

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

Contact. We hypothesized that participants who saw the belief-violating evidence would be more likely to choose actions that violate the principles in the maze games, compared to those who saw belief-consistent evidence or no evidence.

Experiment 1

Method

Participants One hundred and seventy-five adults (mean age = 21.41 years; range = 18 to 52; *SD* = 3.95; 133 women, 37 men, and 5 people were non-binary or did not indicate their gender) participated on an online research platform at a university. Participants provided written informed consent prior to the experiment, and they received 0.5 course credit for a 25-minute experiment.

Stimuli and Procedure Participants were told that they would watch some videos to observe how objects move, and then play some maze games. They were randomly assigned to one of the three conditions, the Baseline condition, the Belief Consistent (BC) condition, and the Belief Violation (BV) condition. They were tested on 3 principles – Solidity, Continuity, and Contact – in counterbalanced orders. For each principle, participants observed 2 types of familiarization trials (3 trials of each type), and then played a maze game with 4 test trials and 2 control trials that were intermixed. In the familiarization trials, participants observed events that were consistent with the principles (BC condition), inconsistent with the principles (BV condition), or did not observe the critical outcome relevant to the principle (Baseline condition).

In the maze games, participants were told their goal was to move the yellow ball from the start location (marked by a green square) to the end location (marked by a red square) as quickly as possible, and they were given a time limit (30, 35, or 45 seconds). Then, they watched videos of the yellow ball moving in the maze, with a timer at the top right corner. For the test trials and the control trials, the video paused to let participants make choices, and resumed after they made their choices. In test trials, participants chose between actions consistent with the principle, *Belief-Consistent (BC) responses*, or actions violating the principle, *Belief-Violation* (*BV*) *responses*. In the control trials, participants made choices that were irrelevant to the target principles. At the end of the game, participants were told that they won or lost (based on whether they finished within the time limit).

At the end of the study, participants in the BC and BV conditions were asked to explain one of the familiarization events for each principle. Then all participants were shown their choices in the last test trial of each principle, and asked to explain why they made their choices.

Solidity principle. In the first type of familiarization trials, a grey ramp, a light blue wall, a dark blue wall, an orange screen, and a ball appeared. The orange screen rotated and occluded the lower half of the walls. The ball rolled down the ramp and moved behind the screen. In the Baseline condition,

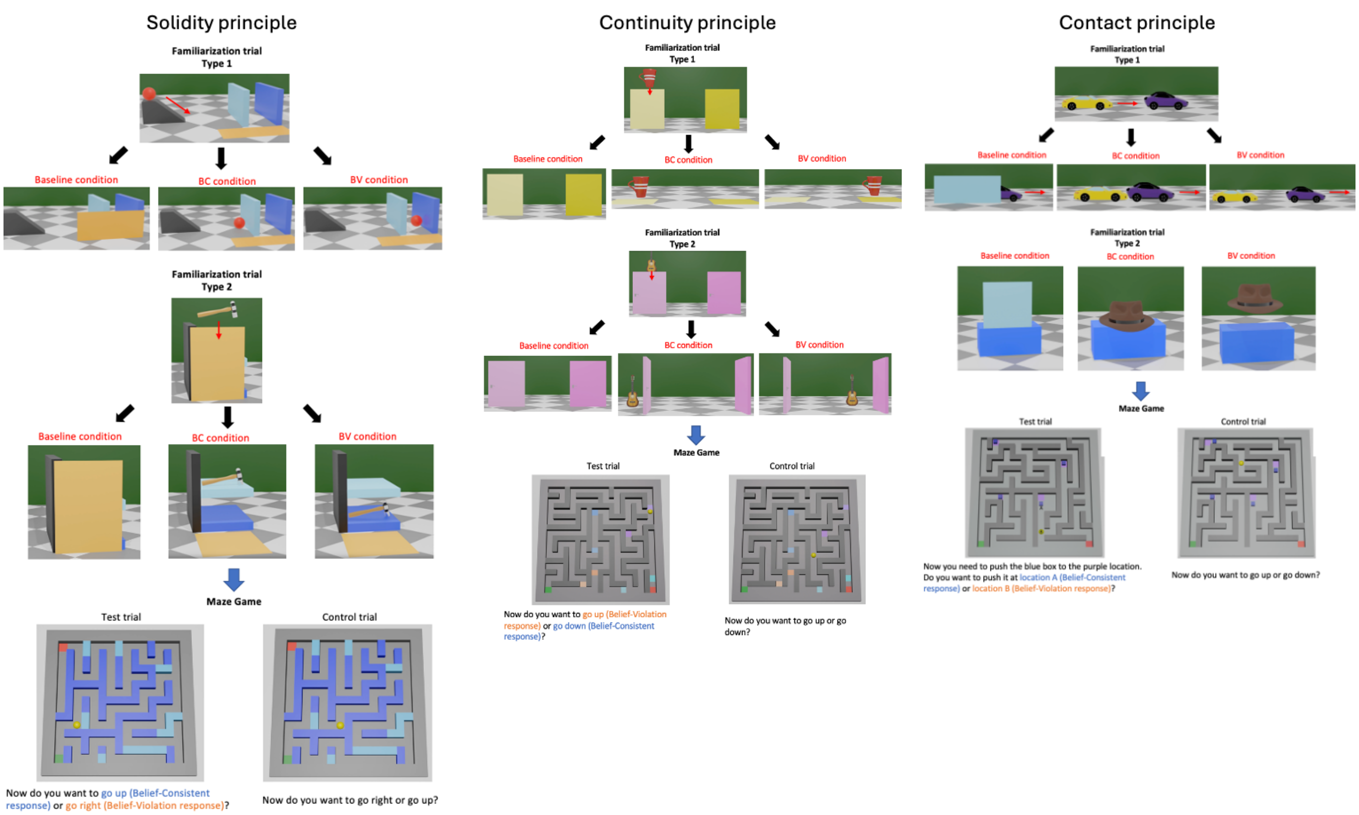

Figure 1: Events shown in the familiarization trials and test trials for the Solidity, Continuity, and Contact principles in Experiment 1.

the screen did not rotate to reveal the location of the ball. In the other two conditions, the screen rotated, and the ball was either before the light blue wall (BC condition) or between the light blue wall and the dark blue wall (BV condition) (Figure 1). A different ball was used in each familiarization trial. The second type of familiarization trials was similar to the first type except the light blue and dark blue walls were placed horizontally, and an object dropped from the top of the screen toward the walls (Figure 1).

The maze was composed of light blue walls that were penetrable and dark blue walls that were not penetrable. In the test trials, the ball stopped beside a light blue wall, where going through the light blue wall would be a shortcut to the end location. Participants were asked which direction they would go. The direction to go through the light blue wall was the *BV response*, and the other direction was the *BC response* (Figure 1). In the control trials, the ball stopped beside a dark blue wall, and participants were asked whether they would go in the direction of the dark blue wall or go in the direction without walls (Figure 1).

At the end of the experiment, participants in the BC and BV conditions were asked to explain why the object appeared at the respective locations in one of the familiarization trials. Then all participants were asked to explain why they chose to go in the respective directions in the last test trial.

Continuity principle. In the first type of familiarization trials, two screens of the same color but different shades appeared. An object dropped from the top and disappeared behind the lighter-shaded screen. In the Baseline condition, the screens did not rotate to reveal the location of the object. In the other two conditions, the screen rotated, and the object was either behind the lighter-shaded screen (BC condition) or behind the darker-shaded screen (BV condition) (Figure 1). A different object was used in each familiarization trial. The second type of familiarization trials was the same as the first type except 2 doors replaced the 2 screens (Figure 1).

The maze was composed of pairs of bricks of the same color but different shades. Objects could teleport from the lighter-shaded bricks to the darker-shaded bricks of the same color. In the test trials, the ball stopped beside a lighter-shaded brick. Participants were asked which direction they would go. The direction of the lighter-shaded bricks was the *BV response*, and the other direction was the *BC response* (Figure 1). In the control trials, the ball stopped at locations that were not beside the colored bricks, and participants were asked which direction they would go (Figure 1).

At the end of the experiment, participants in the BC and BV conditions were asked to explain why the object appeared at the respective locations in one of the familiarization trials. All participants were then asked to explain why they chose to go in the respective directions in the last test trial.

Contact principle. In the first type of familiarization trials, a yellow car launched another car either by contacting it (BC condition) or at a distance (BV condition). In the Baseline condition, a screen blocked the view between the yellow car and the other car (Figure 1). The yellow car launched a different car in each familiarization trial.

In the second type of familiarization trials, a musical toy appeared, and an object was placed either on the toy (BC condition) or above the toy (BV condition), or a screen blocked the location of the object (Baseline condition), and immediately the toy lit up and played music for 5 seconds (Figure 1). A different object was used to activate the toy in each familiarization trial.

In the maze game, in addition to the general instructions to move the yellow ball from the start location to the end location as quickly as possible, participants were also instructed to push 4 blue boxes to the purple locations. In the test trials, the ball stopped at the intersections of the road that led to the end location and the road that led to the blue box. Participants were asked where they would push the box. The location right next to the box was the *BC response*, and the location at a distance was the *BV response*. In the control trials, the ball stopped at intersections without boxes, and participants were asked which direction they would go (Figure 1).

At the end of the experiment, participants in the BC and BV conditions were asked to explain why the yellow car launched the other car in one of the familiarization trials. Then all participants were asked to explain why they chose to push the box at the respective locations in the last test trial.

Results

The proportion of *BV response* by condition and principle is shown in Figure 2. We used mixed-effect logistic regression to predict participants' binary choice (*BV response* = 1, *BC response* = 0) from condition, principle, test trial order, age, gender, and their interactions, while controlling for the random effects of individual participants. The best-fitting model included the interaction of condition and principle and the interaction of principle and test trial order. For the Solidity and the Continuity principles, participants were more likely to choose the *BV response* in the BV condition than in the Baseline condition (Solidity: $\beta = 0.98$, $SE = 0.30$, $p < .001$; Continuity: $\beta = 1.83$, $SE = 0.32$, $p < .001$) and the BC condition (Solidity: $\beta = 0.61$, $SE = 0.29$, $p = .03$; Continuity: $\beta = 1.56$, $SE = 0.30$, $p < .001$); their choices did not differ between the Baseline and the BC conditions (Solidity: $\beta = 0.37$, $SE = 0.30$, $p = .22$; Continuity: $\beta = 0.27$, $SE = 0.33$, $p = .42$). For the Contact principle, participants' choices did not differ between conditions (ps > .2). The interaction of principle and test trial order showed that for the Solidity principle, participants were more likely to choose the *BV response* in the fourth trial than in the first trial ($\beta = 1.52$, $SE = 0.27$, $p < .001$) and the second trial ($\beta = 0.58$, $SE = 0.25$, $p = .02$). For the Contact principle, participants were more likely to choose the *BV response* in the fourth trial than in the first trial ($\beta = 1.46$, $SE = 0.25$, $p < .001$), the second trial ($\beta = 1.74$, $SE = 0.26$, $p < .001$), and the third trial ($\beta = 1.46$, $SE = 0.25$, $p < .001$).

We next analyzed participants' choices in the first trial of each maze. Mixed-effect logistic regression revealed an interaction of condition and principle. For Solidity and Continuity, participants were more likely to choose the *BV response* in BV than in Baseline (Solidity: $\beta = 1.04$, $SE = 0.51$, $p = .04$; Continuity: $\beta = 1.91$, $SE = 0.47$, $p < .001$) and BC conditions (Solidity: $\beta = 0.94$, $SE = 0.49$, $p = .052$; Continuity: $\beta = 1.73$, $SE = 0.44$, $p < .001$); their choices did not differ between Baseline and BC conditions ($ps > .7$). For Contact, their choices did not differ between conditions ($ps > .08$).

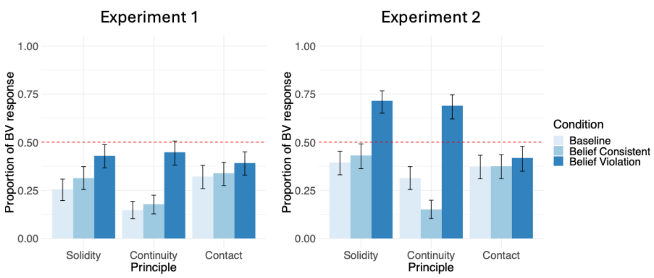

Figure 2: The proportion of trials that participants selected the *BV response* by condition and principle, in Experiments 1 and 2. The dashed line indicates chance selection (.5), and the error bars indicate bootstrapped 95% CIs.

Two researchers coded participants' explanations of the familiarization events into different categories (Cohen's Kappa = .84). In the BC condition, 85% of the responses referred to the principle itself to explain the evidence, 1% referred to the pattern in the evidence, and 14% referred to irrelevant aspects of the events or were incomprehensible. In the BV condition, we categorized participants' explanations based on the criteria in Table 1. Table 1 shows the number and percentage of responses coded within each category for each principle. We used mixed-effects multinomial logistic regression to predict participants' explanation type from principle. We found a significant effect of principle. Participants were less likely to provide "accept evidence" explanations than "explain away" explanations for the Contact principle, compared to the other two principles ($ps < .001$), and they were less likely to provide "accept evidence" explanations than "other" explanations for the Contact principle, compared to the Solidity principle ($p < .001$).

Next, we used mixed-effects logistic regression to predict participants' binary choice in the test trials (BV response = 1, BC response = 0) from the type of explanation they provided. We found that participants were more likely to choose the *BV response* for the principle if they provided "accept evidence" than if they provided "explain away" ($\beta = 0.64$, $SE = 0.25$, $p = .01$) or "other" explanations ($\beta = 0.91$, $SE = 0.45$, $p = .04$).

Lastly, 2 researchers coded participants' explanations of their choices in the maze games into different categories (Cohen's Kappa = .81). For those who chose the *BC response*, 86% of the responses referred to their goal in the maze game, 8% referred to the principles, and 6% referred to irrelevant aspects of the maze or were incomprehensible.

For those who chose the *BV response*, we categorized their explanations based on the criteria in Table 2. Table 2 shows the number of responses coded within each category for each principle. Mixed-effects logistic regression showed that for Contact, participants were less likely to provide "specific rule" than any other types of explanations, compared to Solidity, and less likely to provide "specific rule" than "goal" and "other" explanations, compared to Continuity ($ps < .02$). For Contact, they were also less likely to provide "general rule" than "goal" and "other" explanations, compared to Continuity, and less likely to provide "general rule" than "goal" explanations, compared to Solidity ($ps <$ 001). For Continuity, they were less likely to provide "specific rule" than "general rule" explanations, compared to Solidity ($p < .001$).

Discussion

These results suggest that adults can revise their beliefs about the solidity and continuity principles and generalize their revised beliefs to different behaviors in a different virtual environment given counterevidence. Although there was an order effect for Solidity and Contact, the condition differences were already present in the first trials. This suggests that participants generalized their revised beliefs to the new environment as soon as they entered it; they did not learn to choose

Table 1: Explanations of the familiarization event: Coding criteria and number (percentage) of responses by principle in Experiments 1 and 2

Category	Criterion (and examples for Solidity)	Experiment 1			Experiment 2		
		Solidity	Continuity	Contact	Solidity	Continuity	Contact
Accept Evidence	Accepted the violation of the target principle in the counterevidence (e.g., "The car can go through the light blue wall.").	40 (67%)	33 (55%)	27 (45%)	43 (74%)	43 (74%)	41 (71%)
Explain Away	Explained the counterevidence with reasons that did not involve violations of the target principle (e.g., "There is space between the light blue wall and the screen.").	19 (32%)	19 (32%)	27 (45%)	13 (22%)	10 (17%)	14 (24%)
Pattern	Noted the pattern in the evidence (e.g., "Because it always ends up there.").	0 (0%)	2 (3%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)
Other	Irrelevant or incomprehensible responses (e.g. "I am not sure what happened.").	1 (2%)	6 (10%)	6 (10%)	2 (3%)	4 (7%)	3 (5%)

the principle-inconsistent actions during the maze game.

Participants' explanations of the counterevidence suggest that a majority had accepted the counterevidence, and these participants were indeed more likely to choose actions that violated the principles in the maze games. A minority explained away the counterevidence, and very few participants explicitly appealed to the statistical pattern in the counterevidence. Participants' explanations of their belief-violating choices in the mazes suggest that most participants figured out either the specific rule or the general rule that the principle can be violated in the experiment.

Yet for the contact principle, observing the counterevidence did not change participants' actions in the maze game. Their explanations suggest that participants were more likely to explain away the counterevidence for this principle, for example, "a person drove away the purple car to avoid collision", suggesting that some participants might have thought the objects were controlled by agents. Participants were also less likely to figure out the rule that the contact principle could be violated. In Experiment 2, we examine whether adults would be more likely to revise the contact principle if the connections between the counterevidence and the maze were made more explicit. We also diversified the objects in the counterevidence (i.e., showing that a helicopter, a book, and a die can launch other objects at a distance), so that it would be harder to come up with alternative interpretations.

Experiment 2

Method

Participants One hundred and seventy-five adults (mean age = 21.05 years; range = 18 to 47; *SD* = 3.65; 125 women, 44 men, and 6 people were non-binary or did not indicate their gender) participated on an online research platform at a university. Participants provided written informed consent prior to the experiment, and they received 0.5 course credit for a 25-minute experiment.

Stimuli and Procedure The procedure of Experiment 2 was similar to that of Experiment 1, with a few modifications. First, we added a cover story. We told participants that scientists discovered a new planet; they would observe how objects move and interact on this new planet, and then play a few maze games on this planet. Second, we explicitly told participants that the objects in the mazes were made of the same materials as the objects they observed in the familiarization trials. Third, we made the familiarization trials more diverse by using objects of 3 different categories for each type of familiarization trials for each principle (e.g., we used a ball, a car, and a soda can for the first type of familiarization trials for the Solidity principle, instead of 3 balls in Experiment 1). The rest of the procedure was the same as Expt. 1.

Results

The proportion of *BV response* by condition and principle is shown in Figure 2. Mixed-effect logistic regression revealed effects of age, gender, interaction of condition and principle, and interaction of principle and test trial order. For the Solidity principle, participants were more likely to choose the *BV response* in the BV condition than in the Baseline condition (β = 1.67, *SE* = 0.29, *p* < .001) and the BC condition (β = 1.49, *SE* = 0.29, *p* < .001); their choices did not differ between the Baseline and the BC conditions (β = 0.18, *SE* = 0.27, *p* = .51). For the Continuity principle, participants were more likely to choose the *BV response* in the BV condition than in the Baseline condition (β = 1.86, *SE* = 0.29, *p* < .001) and the BC condition (β = 2.9, 7*SE* = 0.32, *p* < .001); they were less likely to choose the *BV response* in the BC condition than in the Baseline condition (β = -1.10, *SE* = 0.31, *p* < .001). For the Contact principle, participants' choices did not differ between conditions (*p*s > .6).

The interaction of principle and test trial order showed that for the Solidity principle, participants were more likely to choose the *BV response* in the fourth trial than in the first trial (β = 1.38, *SE* = 0.26, *p* < .001). For the Contact principle, participants were more likely to choose the *BV response* in the fourth trial than in the first trial (β = 1.84, *SE* = 0.26, *p* < .001), the second trial (β = 1.32, *SE* = 0.25, *p* < .001), and the third trial (β = 1.38, *SE* = 0.25, *p* < .001).

Table 2: Explanations of choices in mazes: Coding criteria and number (percentage) of responses by principle in Experiments 1 and 2

Category	Criterion (and examples for Continuity)	Experiment 1			Experiment 2		
		Solidity	Continuity	Contact	Solidity	Continuity	Contact
Specific Rule	Referred to the specific rule of how the principle can be violated (e.g., "Light-shaded brick would teleport the ball to dark-shaded brick.").	34 (63%)	14 (52%)	6 (16%)	64 (60%)	31 (45%)	12 (18%)
General Rule	Referred to the general rule that the principle can be violated (e.g., "I can teleport.").	4 (7%)	8 (30%)	1 (3%)	17 (16%)	23 (33%)	4 (6%)
Goal	Referred to the goal in the maze game (e.g., "To get closer to the end location.").	13 (24%)	2 (7%)	25 (68%)	18 (17%)	7 (10%)	45 (66%)
Other	Irrelevant or incomprehensible responses (e.g., "I just guessed.").	3 (6%)	3 (11%)	5 (14%)	7 (7%)	8 (12%)	7 (10%)

Across conditions and principles, participants were less likely to choose the *BV response* with increasing age (β = -0.19, *SE* = 0.09, *p* = .04), and women were more likely to choose the *BV response* than men (β = 0.55, *SE* = 0.21, *p* = .01) and non-binary individuals (β = 1.03, *SE* = 0.51, *p* = .04).

We next analyzed the first trial data. Mixed-effect logistic regression revealed an interaction of condition and principle. For Solidity and Continuity, participants were more likely to choose the *BV response* in BV than in Baseline (Solidity: β = 1.31, *SE* = 0.41, *p* = .001; Continuity: β = 1.60, *SE* = 0.42, *p* < .001) and BC conditions (Solidity: β = 1.29, *SE* = 0.41, *p* = .002; Continuity: β = 2.86, *SE* = 0.50, *p* < .001); for Continuity, participants were less likely to choose the *BV response* in BV than in Baseline condition (β = -1.26, *SE* = 0.48, *p* = .008), for Solidity, their choices did not differ between Baseline and BC conditions (*p* > .9). For Contact, their choices did not differ between conditions (*p*s > .6).

Two researchers coded participants' explanations using the same criteria as in Experiment 1 (Cohen's Kappa = .77). For their explanations of the familiarization events, in the BC condition, 76% of the responses referred to the principle itself to explain the evidence, 1% referred to the pattern in the evidence, and 22% referred to irrelevant aspects of the events or were incomprehensible. In the BV condition, participants' explanation types did not differ across principles (Table 1). Mixed-effects logistic regression showed that participants were more likely to choose the *BV response* in the test trials if they provided "accept evidence" than if they provided "explain away" (β = 0.87, *SE* = 0.26, *p* < .001), "pattern" (β = 1.80, *SE* = 1.38, *p* = .04) or "other" explanations (β = 1.32, *SE* = 0.47, *p* = .005).

For participants' explanations of their choices in the maze games, for those who chose the *BC response*, 78% of the responses referred to their goal in the maze game, 13% referred to the principles, and 9% referred to irrelevant aspects of the maze or were incomprehensible.

For those who chose the *BV response* (Table 2), we found that for Contact, participants were less likely to provide "specific rule" than any other types of explanations, compared to Solidity, and less likely to provide "specific rule" than "goal" and "other" explanations, compared to Continuity (*p*s < .02). For Contact, they were also less likely to provide "general rule" than "goal" and "other" explanations, compared to Continuity and Solidity (*p*s < 001). For Continuity, they were less likely to provide "specific rule" than "general rule" explanations, compared to Solidity (*p* < .001).

Discussion

Experiment 2 replicated the main findings of Experiment 1—adults reliably chose actions violating the solidity and continuity principles after observing belief-violating evidence, but not for the contact principle. We also replicated the order effects for solidity and contact principles.

A majority of participants accepted the counterevidence, even for the contact principle, and these participants were more likely to choose actions that violated the principles in the mazes. A minority explained away the counterevidence, and very few learned from the statistical pattern.

General Discussion

The present study investigated the extent to which adults would generalize their revised beliefs after observing counterevidence about core object principles. We found that given a diverse set of counterevidence, adults revised their beliefs about the solidity and continuity principles, and they generalized these revised beliefs to guide their behaviors in a different virtual environment.

Compared to past studies (Liu & Xu, 2022, under review), more participants accepted the counterevidence, and very few relied on the statistical pattern of the counterevidence in the present study. One possible reason is that we used a more diverse set of counterevidence (i.e., 2 types of familiarization events and multiple categories of objects), making it less likely for participants to simply notice the pattern in the events and more likely to consciously process the violations in the counterevidence. In addition, a majority of participants were able to explicitly articulate their revised beliefs about the principles and use these revised beliefs to explain their belief-violating actions in the maze games.

Lastly, counterevidence about the contact principle did not reliably change adults' behaviors in the subsequent maze game. Why? One possibility is that for the contact principle, the violations in the counterevidence and the choices in the maze game were the least similar (Figure 1) among all 3 principles. Another possibility is that participants had to simultaneously achieve two goals in the maze game for the contact principle – getting to the end location as quickly as possible and pushing all 4 boxes into the corresponding locations – which may have increased the computational cost required for this game compared to the games for the other two principles. In future studies, we will make the generalization task more similar to the counterevidence and reduce the computational cost of the task to examine if adults would also generalize their revised beliefs about the contact principle to different environments.

In conclusion, the present research provides preliminary evidence that the core knowledge principles about objects are not completely encapsulated. Adults can learn from violations of these principles in a virtual environment, and use the revised principles to guide their actions in a new virtual environment. Future studies should examine whether adults can learn from violations of these principles in the real world and generalize their revised beliefs to different real-world environments.

References

Aguiar, A., & Baillargeon, R. (1999). 2.5-Month-Old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*, *39*(2), 116–157.

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of* *the National Academy of Sciences*, *117*(47), 29302–29310.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*(1), 2-13.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., Bolz, M., Henrich, J., Setoh, P., Wang, J., & Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755), 20122654.

Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press, USA.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108.

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science, 306*, 496 – 499.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and Blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35*(8), 1407–1455.

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition*, *157*, 61–76.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal behaviour*, *61*(1), 139-151.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The Naïve Utility Calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288.

Liu, R. & Xu, F. (2021). Revising core beliefs in young children. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar (Eds.), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.* Cognitive Science Society.

Liu, R. & Xu, F. (2022). Can adults revise their core beliefs about objects? In J. Culberston, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Liu, R. & Xu, F. (under review). Children and adults revise their core principles about objects and agents with statistical evidence.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using Hierarchical Bayesian Models. *Cognitive Science, 34*(1), 113–147.

Piaget, J. (1954). *The construction of reality in the child.* New York, NY: Routledge.

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, *57*(7), 971–976.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80*(1–2), 159–177.

Spelke, E. S. (1988). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language*. Clarendon Press/Oxford University Press.

Spelke, E. S. (2000). Core knowledge. *American Psychologist, 55*(11), 1233–1243.

Spelke, E. S. (2022). *What babies know: Core knowledge and composition volume 1* (Vol. 1). Oxford University Press.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological review*, *99*(4), 605.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*, 533-558.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14*(5), 498–504.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841–864.