The Development of Weight-Based Inductive Reasoning in Preschool Children: A China-US Cross-Cultural Comparison (Post-Print)

Authors: Wang Zhidan, Zhou Aibao, Ronghua Zhang, Bu Mengjin, Li Yuwen, Haijing Wang, Williamson Rebecca

Date: 2018-09-07T00:00:00+00:00

Abstract

Research on object weight cognition constitutes one of the important approaches to investigating children' s overall cognitive development. The present study examined the developmental characteristics and cross-cultural consistency of inductive reasoning based on weight utilizing linguistic labels in American and Chinese children through two experiments. Experimenters employed linguistic labels to cue one of three objects with completely identical external perceptual features, subsequently requiring children to select from the remaining two objects the one they believed shared the linguistic label with the target object. The study investigated whether children could make categorical judgments based on the intrinsic property of object weight. The results demonstrated: (1) Inductive reasoning based on weight among 4- and 5-year-olds was significantly higher than that of 3-year-olds, with no significant difference between 4- and 5-year-olds; (2) The performance of both American and Chinese 4- and 5-year-olds was significantly above chance level, whereas 3-year-olds did not differ significantly from chance level; (3) American 2-year-olds' performance was marginally below chance level, while 6-year-olds performed significantly above chance level. In summary, children' s inductive reasoning based on weight using linguistic labels develops with age; American and Chinese children exhibit cross-cultural consistency in developmental characteristics and age transitions, and 4-year-olds have acquired the ability to utilize linguistic labels for inductive reasoning based on object intrinsic properties.

Full Text

The Development of Preschool Children' s Inductive Reasoning About Weight: A Cross-Cultural Comparison of Chinese and American Children

WANG Zhidan[1], **ZHOU Aibao**[2], **ZHANG Ronghua**[2], **BU Mengjin**[1], **LI Yuwen**[1], **WANG Haijing**[1], **WILLIAMSON Rebecca**[3]

[1] School of Education Science, Jiangsu Normal University, Xuzhou 221116, China

[2] School of Psychology, Northwest Normal University, Lanzhou 730070, China

[3] Department of Psychology, Georgia State University, Atlanta 30302, USA

Abstract

Research on weight cognition represents an important pathway for investigating children' s cognitive development. This paper examines the developmental characteristics and cross-cultural consistency of American and Chinese children' s use of verbal labels to make inductive inferences based on weight through two experiments. Experimenters used verbal labels to cue one of three objects with completely identical external perceptual features, then asked children to select from the remaining two objects the one they believed shared the same label as the target object. The study investigated whether children could make categorical judgments based on the internal property of weight. Results showed: (1) Four- and five-year-olds performed significantly better than three-year-olds in using weight for inductive reasoning, with no significant difference between four- and five-year-olds; (2) Both American and Chinese four- and five-year-olds performed significantly above chance level, while three-year-olds did not differ from chance; (3) American two-year-olds performed marginally below chance level, while six-year-olds performed significantly above chance. In summary, children' s ability to use verbal labels for weight-based inductive reasoning develops with age, and American and Chinese children show cross-cultural consistency in developmental patterns and age-related transitions, with four-year-olds demonstrating the capacity to use verbal labels for inductive reasoning based on object internal properties.

Keywords: weight; inductive reasoning; internal property; cross-cultural consistency
Classification Number: B844

1 Introduction

The fundamental core of scientific thinking is scientific abstraction. Phenomena are typically external and observable, whereas essence is internal and cannot be directly perceived with the naked eye. To understand the nature and laws of things, one must rely on abstract thinking to reveal the hidden essence behind them. For instance, after observing the phenomenon of an apple falling, Newton discovered and formulated the law of universal gravitation. In this example, the falling apple represents an external phenomenon, while gravity represents the internal essence. An important indicator of children' s cognitive development is their ability to reveal the essence of things through their phenomena (Gelman & Wellman, 1991; Wellman & Gelman, 1988). Research in social cognition has shown that by age five, most children have basically developed theory-of-mind abilities, enabling them to understand others' mental states (such as internal beliefs, desires, and intentions) to predict their behavioral responses (Flavell, Green, Flavell, Harris, & Astington, 1995; Liu, Wellman, Tardif, & Sabbagh, 2008). However, relatively few studies have directly investigated children' s understanding of internal properties in the physical cognition domain. Therefore, this research uses children' s cognition of the internal property of weight as an entry point and adopts a cross-cultural approach to examine the developmental characteristics of inductive reasoning in American and Chinese children.

Compared to basic visual attributes (such as color and shape), weight is an internal property of objects. Although weight can be preliminarily predicted based on object size, it cannot be directly seen. To perceive an object' s weight, individuals typically need direct physical contact with it. Therefore, weight provides an excellent concrete example for studying whether children can engage in inductive reasoning based on object internal properties (Povinelli, 2012; Wang, Williamson, & Meltzoff, 2018). The weight attribute of objects has long been a focus for developmental psychologists studying children's cognitive development, particularly yielding substantial findings in children' s conservation and rule application (e.g., Piaget, 1952; Siegler & Chen, 1998). Piaget (1952) argued that preoperational children cannot conserve weight and can only understand weight based on object size and shape. Only after age seven, during the concrete operational stage, do children gradually recognize that object weight does not change with size and shape. In another classic balance-scale task designed by Piaget, children needed to determine which side of the balance would tilt downward when objects were placed at different positions. He believed that only concrete operational children could solve such problems. However, recent research indicates that even infants possess some understanding of object weight, as they can infer weight magnitude based on external properties such as material and color (Hauf, Paulus, & Baillargeon, 2012; Hauf & Paulus, 2011; Patnaik, 2008; Paulus & Hauf, 2011).

Existing research has primarily examined preschool children' s understanding of object weight through two types of tasks. The first is weight-based causal reasoning tasks, where weight functions as a variable that produces external effects on other objects (Povinelli, 2012; Wang et al., 2018). Studies show that although young preschool children cannot solve the complex problems set by Piaget (1952), they can solve some simple balance-scale tasks (Li, Xie, Yang, & Cao, 2017; Patnaik, 2008; Schrauf, Call, & Pauen, 2011; Siegler & Chen, 1998). In Schrauf et al. (2011), experimenters selected the heavier of two identical-looking objects to place on one side of a balance scale to obtain a hidden target object on the other side. Results showed that four-year-olds chose the heavy object significantly more often than the light object to obtain the target. Another study investigated how children understand weight through object collisions. In Povinelli (2012), children observed experimenters demonstrate interactions between two objects and a target object. When a heavy object was launched from the top of a platform, the experimenter could obtain the target object knocked off a ramp. However, when a light object was launched, the target object became stuck and could not be obtained. Results showed that four-year-olds chose the heavy object significantly above chance level to strike the target, whereas three-year-olds' performance did not differ from chance.

The second type is weight classification tasks, where weight serves as a dimension for distinguishing object categories (Povinelli, 2012; Wang, 2014; Wang, Meltzoff, & Williamson, 2015; Wang, Williamson, & Meltzoff, 2015). For example, in classification tasks, one category consists of heavy objects and another of light objects. Wang et al. (2015) tested whether children could classify objects with identical appearances by weight. In the experimental group, the experimenter picked up an object, performed a weighing motion, then placed it in a tray, subsequently weighing the remaining three objects and placing them in the tray one by one. As a result, two heavy objects were placed in one tray and two light objects in another tray. In Control Group 1, the experimenter only demonstrated weighing the four objects without classifying them. In Control Group 2, the experimenter placed the four objects in trays beforehand, then demonstrated weighing each object and returning it to the tray. During testing, the experimenter used not only the demonstrated objects but also another set of objects with different shapes, colors, and weights to test children' s transfer. Results showed that four-year-olds in the experimental group classified objects by weight significantly more often than both control groups and significantly above chance level, indicating that four-year-olds abstracted the weight classification rule and transferred it to new contexts.

To date, no research has examined children' s inductive reasoning abilities based on the internal property of weight. Therefore, the first purpose of this study is to investigate children' s ability to use verbal labels for weight-based inductive reasoning. Inductive reasoning is an important type of scientific reasoning. Category-based inductive reasoning helps children explain new situations and infer implicit information that has not been explicitly taught (Chen, Wei, Zhong, & Mo, 2014; Noles & Gelman, 2012). During category learning, children extract abstract rules that can also be applied to broader contexts. Thus, studying children' s categorical inductive reasoning helps children explore the unknown and conserve learning resources. Research shows that verbal labels are effective cues that promote inductive reasoning, serving as a vehicle for studying concepts and category knowledge (Althaus & Westermann, 2016; Johansen, Savage, Fouquet, & Shanks, 2015). Wang et al. (2014) found that verbal labels can assist children in category-based inductive reasoning. When two objects share the same verbal label, children believe they should be similar in some respects (Fulkerson, Waxman, & Seymour, 2006). Conversely, when two objects do not share the same label, children actively seek differences between them to explain why the experimenter used different labels. If objects' external features are perceptually similar and cannot directly explain the different labels provided by the experimenter, children will further search for differences in internal properties (Kemler-Nelson, Herron, & Holt, 2003). However, most research in this area has used verbal labels to cue visual external properties of objects, while studies using labels to cue internal properties remain relatively scarce.

This study employs the classic match-to-sample task (Hochmann, Mody, & Carey, 2016; Hochmann et al., 2017) to examine children' s ability to use verbal labels for inductive reasoning based on the internal property of weight. Each set of three objects has similar external perceptual features, with only the internal property of weight distinguishing them. The experimenter selects a target object and labels it with a verbal label. Children are then asked to choose from the remaining two objects the one they believe shares the same label as the target object. If children make selections based on external perceptual similarity between objects, they should choose the remaining two objects with equal probability, as both share identical external features with the target object. From external perceptual similarity, all three objects should belong to the same category. However, if children' s categorical judgments go beyond external perceptual similarity, they may further search for objects' internal properties to distinguish them, thereby selecting the object with the same internal weight as the target object cued by the experimenter' s verbal label.

The second purpose of this study is to explore cross-cultural consistency in children' s weight-based inductive reasoning. Cross-cultural cognitive development research is still in its infancy, and the problem of sample homogeneity is particularly serious in developmental psychology (Nielsen & Haun, 2016; Nielsen, Haun, Kärtner, & Legare, 2017). One analysis of papers published in core developmental psychology journals from 2006 to 2010 found that 90.52% of studies used samples from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (Nielsen et al., 2017). Therefore, conducting cross-cultural cognitive development research has important theoretical value. This paper selects the United States, representing Western cultural backgrounds, and China, representing Eastern cultural backgrounds, to examine cross-cultural consistency in the development of concept-based inductive reasoning using the weight attribute as an example.

Experiment 1 examined the developmental characteristics of American children aged 2-6 using verbal labels for inductive reasoning. Based on Experiment 1' s results, Experiment 2 examined whether Chinese children show developmental patterns and age transitions at the same stage. Research indicates cognitive style differences between American and Chinese children: Chinese children are more adept at relational-contextual reasoning, where they make judgments based on relationships between objects (e.g., a baby and mother belong together because mothers care for babies), whereas American children are more adept at inferential-categorical reasoning, where they make categorical judgments based on object properties (e.g., a boat and jeep belong together because they both have engines as parts) (Chiu, 1972; Richland, Chan, Morrison, & Au, 2010). Since the current experimental task involves categorical judgment based on object internal properties, American children might outperform Chinese children of the same age. However, children' s cognition of weight properties likely depends on direct perceptual experiences with surrounding objects, so weight-based inductive reasoning may be less influenced by cultural and social factors, potentially resulting in no performance differences between American and Chinese children.

Experiment 1

2.1 Participants

One hundred children from Atlanta, USA participated in the experiment. There were 20 children in each age group: 2-year-olds (10 boys, 10 girls), 3-year-olds (11 boys, 9 girls), 4-year-olds (13 boys, 7 girls), 5-year-olds (9 boys, 11 girls), and 6-year-olds (10 boys, 10 girls). Based on pilot study results, G*Power (Erdfelder, Faul, & Buchner, 1996) was used to calculate sample size, with power set at 0.85, alpha at 0.05, and effect size at 0.20, yielding a required sample of 100 participants (20 per age group). According to parent or guardian reports, the sample was 84% White, 6% Asian, 4% African American, 2% Latino, 3% other, and 1% unreported. Researchers recruited and tested children at the Fernbank Natural History Museum in Atlanta. The experimental procedure was approved by the Georgia State University Institutional Review Board, and informed consent was obtained from parents or guardians of all participating children.

2.2 Materials

The experiment included two sets of materials, each containing three objects with completely identical external features–same color, size, and shape. Set A consisted of three ice cream cups, with two heavy ones weighing 94.41g and one light one weighing 15.35g, or one heavy one weighing 94.41g and two light ones weighing 15.35g. Set B consisted of three white bottles, with two heavy ones weighing 147.20g and one light one weighing 6.24g, or one heavy one weighing 147.20g and two light ones weighing 6.24g (Figure 1 [Figure 1: see original paper]). Three adults unaware of the experimental purpose reported significant differences between heavy and light objects.

2.3 Procedure

Prior to the formal experiment, experimenters obtained verbal consent from children. The experimenter said, "Hello, little friend! Let' s play a game together today, would you like to play?" If the child gave an affirmative answer, the procedure continued. If the child shook their head or said no, the procedure was terminated, but the child still received a small gift. After obtaining consent, the formal experiment began. The experiment consisted of two phases:

a familiarization phase and a test phase, with all children participating in both phases.

2.3.1 Familiarization Phase The experimenter selected one set of materials and placed the three objects in a row on the table (e.g., Set A ice cream cups). The experimenter encouraged the child to manipulate each ice cream cup. If the child showed no willingness, the experimenter picked up each cup and handed it to the child until the child became familiar with all objects. This process ensured children formed perceptual experiences of each object and lasted approximately 30 seconds.

2.3.2 Test Phase Immediately following familiarization, formal testing began. The experimenter rearranged the three objects near their side of the table. The experimenter picked up one ice cream cup and said, "Look, this is a dax," then handed this cup to the child. After the child manipulated the object, the experimenter said, "Now can you give the dax back to me?" After the child returned the object, the experimeter placed it back in its original position, picked up the remaining two objects, and said, "There is another dax among these two." Simultaneously, the experimenter handed both objects to the child, asking them to hold one in each hand, and said, "Which one do you think is the dax? Please find it and give it to me."After speaking, the experimenter extended their right hand and waited for the child' s selection. The experimenter' s right hand was placed directly in front of the child, midway between the child' s two hands, to avoid being closer to either object and providing additional cues.

After the child selected and handed an object to the experimenter, the experimenter removed the first set of three ice cream cups, took out Set B' s three white bottles, placed them on the table, and began the second test. The testing procedure was identical to the first set, except that when the experimenter picked up a white bottle, they said, "Look, this is a wug." After the child manipulated the object, the experimenter returned it to its original position, then picked up the remaining two white bottles and said, "There is another wug among these two," while handing both objects to the child and asking, "Which one do you think is the wug? Please find it and give it to me." A different verbal label was used in the second test.

The order of the two material sets (ice cream cups/white bottles) was counterbalanced using an ABBA design. Different material sets corresponded to specific verbal labels—Set A ice cream cups corresponded to the label "dax," and Set B white bottles corresponded to the label "wug."

2.3.3 Scoring The dependent variable was children' s scores for selecting objects matching the target' s weight. If the experimenter picked up a heavy object and the child chose the heavy one among the remaining two objects, the child received a score of 1 for that test; otherwise, they received 0. If the experimenter picked up a light object, the child received 1 point for choosing the light object and 0 for choosing the heavy one. With two tests administered, total target-weight matching scores ranged from 0 to 2. The experimenter recorded children' s responses directly, while an assistant randomly recorded responses for 20% of participants. Inter-rater reliability between the two experimenters was assessed using Cohen' s kappa coefficient (0.95), indicating excellent agreement.

2.4 Results

Analyses indicated that neither participant gender ($p = 0.20$) nor material presentation order ($p = 0.16$) had significant effects on the dependent variable; therefore, these variables were not examined further in subsequent analyses. Kruskal-Wallis tests revealed a significant age effect, $H(4) = 41.75$, $p < 0.001$. Mann-Whitney tests for further analysis showed that the 6-year-old group performed significantly higher than the 2-year-old group ($U = 26$, $r = 0.82$), 3-year-old group ($U = 69$, $r = 0.66$), and 4-year-old group ($U = 129$, $r = 0.42$). The 5-year-old group performed significantly higher than the 2-year-old group ($U = 47$, $r = 0.71$) and 3-year-old group ($U = 94$, $r = 0.52$). The 4-year-old group performed significantly higher than the 2-year-old group ($U = 86$, $r = 0.52$). However, no significant differences were found between the 5- and 6-year-old groups ($U = 180$, $r = 0.17$), between the 4- and 5-year-old groups ($U = 151$, $r = 0.27$), or between the 3- and 4-year-old groups ($U = 146$, $r = 0.26$) (Table 1). These results indicate that American children' s ability to use verbal labels for weight-based inductive reasoning improves with age from 2 to 6 years. One-way ANOVA yielded consistent results with the nonparametric tests.

Table 1 Number and Percentage of American Children Aged 2-6 Selecting Correct Objects by Number of Correct Weight Matches

Age Group	0 correct	1 correct	2 correct
2-year-olds	8 (40%)	10 (50%)	2 (10%)
3-year-olds	2 (10%)	12 (60%)	6 (30%)
4-year-olds	2 (10%)	6 (30%)	12 (60%)
5-year-olds	1 (5%)	2 (10%)	17 (85%)
6-year-olds	0 (0%)	1 (5%)	19 (95%)

With only one correct matching object between the two options in each test, random selection would yield a 50% probability of choosing correctly in a single test. With two tests administered, chance performance was set at 1. One-sample Wilcoxon signed-rank tests compared children' s correct selections to chance level. Results indicated that 4-year-olds ($p = 0.004$), 5-year-olds ($p < 0.001$), and 6-year-olds ($p < 0.001$) selected correct objects significantly above chance. However, 3-year-olds' performance ($p = 0.16$) did not differ significantly from chance, while 2-year-olds' performance was marginally significantly below chance ($p = 0.06$). Two related-samples Wilcoxon signed-rank analysis examined whether significant differences existed between the first and second tests, revealing no significant difference, z = -1.26, p = 0.21, r = 0.20, indicating no learning effects during testing.

In summary, American children' s weight-based inductive reasoning abilities develop with age from 2 to 6 years, with four-year-olds beginning to use verbal labels for weight-based inductive reasoning, and this ability reaching maturity by age six.

Experiment 2

Experiment 2 further examined the developmental characteristics of Chinese children' s use of verbal labels for weight-based inductive reasoning. American two-year-olds showed very low response rates in the experimental task, with 50% essentially choosing randomly and 40% consistently selecting objects opposite to the target' s weight. Therefore, two-year-olds were not tested in Experiment 2. Six-year-olds achieved 95% accuracy in selecting correct objects with no significant difference from five-year-olds, so this age group was also not tested in Experiment 2.

3.1 Participants

Sixty children were randomly selected from two kindergartens in a central Chinese city, including 32 boys and 28 girls. Children were divided into three age groups: 20 three-year-olds (10 boys, mean age = 42.40 months, SD = 2.78), 20 four-year-olds (10 boys, mean age = 55.45 months, SD = 2.33), and 20 five-year-olds (12 boys, mean age = 66.55 months, SD = 3.33). Sample size was determined based on the number of children tested per age group in Experiment 1.

3.2 Materials and Procedure

Experiment 2 was conducted six months after Experiment 1 data collection. All procedures were identical to Experiment 1, with two modifications: First, children were tested in their kindergarten setting. Before testing, experimenters arranged a quiet classroom in the kindergarten as the testing environment, with the entire procedure completed in this room. Second, the verbal labels used for American children were replaced. Researchers created two new labels— "delu" and "peru" —based on Chinese language characteristics, as directly translating English nonsense syllables into Chinese would no longer sound like nonsense syllables. Children' s responses were recorded simultaneously by the experimenter and assistant, with inter-rater reliability kappa coefficient of 0.98, indicating excellent agreement.

3.3 Results

Analyses indicated that neither participant gender (p = 0.90) nor object presentation order (p = 0.52) had significant effects on the dependent variable;

therefore, these variables were not examined further. Kruskal-Wallis nonparametric tests revealed a significant age effect, $H(2) = 18.71$, $p < 0.001$. Mann-Whitney tests further showed that the 5-year-old group performed significantly higher than the 3-year-old group ($U = 68$, $r = 0.61$), and the 4-year-old group performed significantly higher than the 3-year-old group ($U = 85$, $r = 0.54$). However, no significant difference existed between the 4- and 5-year-old groups ($U = 165$, $r = 0.18$) (Table 2). These results indicate that Chinese children' s ability to use verbal labels for weight-based inductive reasoning also improves with age from 3 to 5 years. One-way ANOVA yielded results consistent with the nonparametric tests.

Table 2 Number and Percentage of Chinese Children Aged 3–5 Selecting Correct Objects by Number of Correct Weight Matches

Age Group	0 correct	1 correct	2 correct
3-year-olds	7 (35%)	11 (55%)	2 (10%)
4-year-olds	0 (0%)	5 (25%)	15 (75%)
5-year-olds	1 (5%)	2 (10%)	17 (85%)

One-sample Wilcoxon signed-rank tests compared children' s correct selections to chance level. Results indicated that 4-year-olds ($p < 0.001$) and 5-year-olds ($p < 0.001$) selected correct objects significantly above chance. However, 3-year-olds' performance did not differ significantly from chance ($p = 0.10$). These findings are consistent with American children, showing that Chinese children can also use verbal labels for weight-based inductive reasoning by age four.

A loglinear model analyzed the relationship among culture, age, and target-weight matching scores. The model likelihood ratio was $^2(0) = 0$, $p = 1.00$. Further analysis revealed no significant three-way interaction among culture, age, and target-weight matching scores, $^2(4) = 6.98$, $p = 0.14$, no significant interaction between age and culture, $^2(2) = 0.87$, $p = 0.65$, and no significant interaction between culture and target-weight matching scores, $^2(2) = 3.69$, $p = 0.16$. This indicates that culture does not affect children' s weight-based inductive reasoning. However, the interaction between age and target-weight matching scores was significant, $^2(4) = 31.57$, $p < 0.001$. Mann-Whitney tests showed that the 5-year-old group performed significantly higher than the 3-year-old group ($U = 332$, $r = 0.55$), and the 4-year-old group performed significantly higher than the 3-year-old group ($U = 471$, $r = 0.39$). However, no significant difference existed between the 5- and 4-year-old groups ($U = 629$, $r = 0.22$). In summary, the developmental characteristics of children's weight-based inductive reasoning are consistent with results from separate analyses of American and Chinese children (Figure 2 [Figure 2: see original paper]).

Figure 2 [Figure 2: see original paper] American and Chinese 3- to 5-year-old children' s scores for selecting weight-matching objects

4 General Discussion

The first research question addressed when children begin to use verbal labels for weight-based inductive reasoning. Results showed that 4- and 5-year-olds performed significantly better than 3-year-olds, with no significant difference between 4- and 5-year-olds. Additionally, 4- and 5-year-olds selected correct objects significantly above chance level, while 3-year-olds did not differ from chance. Therefore, we conclude that age four represents a turning point in the ability to use verbal labels for weight-based inductive reasoning. This is consistent with the age transition found in previous research on children' s use of verbal labels for inductive reasoning (Long, Wu, Li, Chen, Feng, & Li, 2006; Gelman & Markman, 1986; Kalish & Lawson, 2007; Kemler-Nelson et al., 2003; Kemler-Nelson, Russell, Duke, & Jones, 2000; Waxman & Klibanoff, 2000).

The second research question concerned whether children' s weight-based inductive reasoning using verbal labels shows cross-cultural consistency. Since American and Chinese children show identical developmental trajectories in weight-based inductive reasoning abilities, this study concludes that these abilities exhibit cross-cultural consistency. This result also aligns with Wang et al. (2015), who found that American and Chinese four-year-olds learned to categorize objects by weight through observing others. The absence of cross-cultural differences may be because weight is a physical property that focuses on children' s understanding of objective natural phenomena, which children in both cultural backgrounds can acquire through perceptual contact with surrounding objects, thus being less influenced by cultural factors. Of course, this study does not argue that cultural background never affects children' s inductive reasoning; it is also possible that children in this study were relatively young, and cultural and social factors had not yet established stable influences on their inductive reasoning.

The study found that age four is the developmental turning point for children using verbal labels for weight-based inductive reasoning. Previous research also confirms that four-year-olds, but not three-year-olds, more significantly use the heavier of two identical-looking objects to tip a balance beam (Schrauf et al., 2011), use heavy objects to slide down a ramp to move another object (Povinelli, 2012), and categorize different objects by weight (Wang et al., 2015). This study argues that three-year-olds' inability to engage in weight-based inductive reasoning does not stem from misunderstanding the experimental task, as research shows that three-year-olds can comprehend match-to-sample tasks (Gelman & Wellman, 1991), and some studies even indicate that infants around 14 months can understand match-to-sample tasks (Graham, Kilbreath, & Welder, 2004; Welder & Graham, 2001; Hochmann et al., 2016). Therefore, this study proposes that three-year-olds lack weight-based inductive reasoning abilities. What abilities develop significantly between ages three and four? This study suggests three possibilities.

First, four-year-olds' abstract thinking about object internal properties develops significantly. Three-year-olds rely more on external properties to distinguish objects, whereas four-year-olds have greater awareness of object internal properties and can differentiate objects based on internal features and functions (Buchanan & Sobel, 2011; Erb, Buchanan, & Sobel, 2013; Nazzi & Gopnik, 2000). In Sobel et al. (2007), children saw two objects with different internal structures, one of which could make a target object light up while the other could not. When later judging which objects could make the target light up, only four-year-olds recognized that a specific internal structure was necessary for the target to light up. In this study, all objects had identical external features, so children could not distinguish them based on external perceptual similarity. In this situation, three-year-olds might consider all objects to belong to the same category, resulting in random selections. However, when four-year-olds cannot distinguish objects through external features, they realize that objects differ in the internal property of weight, which can serve as the determining factor for distinguishing objects, leading them to select the object with the same weight as the target.

Second, four-year-olds' inhibitory control abilities develop significantly. Three-year-olds may also be able to consider object internal properties, as some research suggests that even infants can abstractly represent object weight (e.g., Hauf et al., 2012; Hauf & Paulus, 2011; Paulus & Hauf, 2011). However, they may still lack the inhibitory control needed to use this internal property when external properties are identical. In this study, the three objects' appearances were completely identical, so from perceptual similarity, all three objects should belong to the same category. To differentiate objects by weight, children must inhibit the fact that the three objects appear identical externally and instead consider their internal property differences. That is, children must overcome perceptual similarity of external attributes while examining differences in internal attributes, which remains quite difficult for three-year-olds. Consistent with this, in classic appearance-reality theory-of-mind tasks, three-year-olds still relatively lack the ability to inhibit external attributes when they conflict with internal attributes.

Third, four-year-olds' language expression abilities develop significantly. Language development may greatly promote four-year-olds' use of verbal labels for weight-based inductive reasoning (Schrauf et al., 2011). With age, four-year-olds more spontaneously use vocabulary such as "heavy" and "light" to describe objects in experimental tasks. Children' s use of such vocabulary may represent a self-explanation process. Research shows that, on one hand, self-explanation can help children focus more attention on core elements relevant to task solution, enabling them to ignore irrelevant elements when solving tasks (Legare, 2012). On the other hand, similar verbal expressions may assist children' s thinking processes, making them more likely to search for internal causes to explain external phenomena (Legare, 2014; Walker, Lombrozo, Legare, & Gopnik, 2014).

This study was conducted in two countries where testing environments could not be kept completely identical. However, researchers ensured that testing procedures, materials, and experimenters were identical, which to some extent ensured the validity of cross-cultural comparisons. Although this study did not collect background information such as family socioeconomic status, museum data indicated that most visitors were local middle-class families, and families registered at the Chinese kindergartens were also primarily middle-class. Therefore, the American and Chinese children tested were relatively comparable. This study' s innovations include: First, it examines cross-cultural consistency in weight-based inductive reasoning between the United States (representing Western culture) and China (representing Eastern culture), which has important theoretical significance for better understanding cognitive development across cultural backgrounds. Second, it selects the internal property of weight as an entry point for examining children' s inductive reasoning abilities, which is innovative compared to previous research focusing primarily on visual external properties. Third, it identifies age four as the developmental turning point for weight-based inductive reasoning, which can help identify critical periods for better inductive reasoning training and improve children' s scientific literacy.

Future research can continue in two directions: First, this study only selected weight as a specific case for examining children' s inductive reasoning based on object internal properties; future research could select other internal properties (e.g., sound) to examine whether inductive reasoning based on different internal properties shows consistent developmental patterns. Second, age four is a turning point both for theory-of-mind in social cognition (e.g., Wellman, Fang, Liu, Zhu, & Liu, 2006; Liu et al., 2008) and for reasoning about object internal properties in physical cognition (e.g., Schrauf et al., 2011; Wang et al., 2018). Future research could explore whether the internal mechanisms underlying social cognitive reasoning and physical cognitive reasoning tasks are consistent.

5 Conclusions

This study draws the following main conclusions: 1. Children can use verbal labels for weight-based inductive reasoning, with age four being the developmental turning point for this ability. 2. American and Chinese children show cross-cultural consistency in developmental characteristics and age transitions for weight-based inductive reasoning.

References

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, 151, 5−17.

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, 82(6), 2053−2066.

Chen, L., Wei, X. M., Zhong, L. J., & Mo, L. (2014). The role of causal relations between new property and prototype in induction. *Psychological Exploration*, 34(5), 398−404. [陈琳, 魏晓玛, 钟罗金, 莫雷. (2014). 基于原型特征因果关系的归纳推理研

究. 心理学探新, 34(5), 398−404.]

Chiu, L-H. (1972). A cross-cultural comparison of cognitive styles in Chinese and American children. *International Journal of Psychology*, 7(4), 235−242.

Erb, C. D., Buchanan, D. W., & Sobel, D. M. (2013). Children' s developing understanding of the relation between variable causal efficacy and mechanistic complexity. *Cognition*, 129(3), 494-500.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28(1), 1−11.

Flavell, J. H., Green, F. L., Flavell, E. R., Harris, P. L., & Astington, J. W. (1995). Young children' s knowledge about thinking. *Monographs of the Society for Research in Child Development*, 60(1), 1−96.

Fulkerson, A. L., Waxman, S. R., & Seymour, J. M. (2006). Linking object names and object categories: Words (but not tones) facilitate object categorization in 6- and 12-month-olds. In Bamman, D., Magnitskaia, T., & Zaller, C. (Eds.) *Supplement to the Proceedings of the 30th Boston University Conference on Language Development*. Cascadilla Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23(3), 183−209.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, 38(3), 213−244.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, 75(2), 409−427.

Hauf, P., & Paulus, M. (2011). Experience matters: 11-month-old infants can learn to use material information to predict the weight of novel objects. *Infant Behavior and Development*, 34(3), 467−471.

Hauf, P., Paulus, M., & Baillargeon, R. (2012). Infants use compression information to infer objects' weights: examining cognition, exploration, and prospective action in a preferential-reaching task. *Child Development*, 83(6), 1978−1995.

Hochmann, J-R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match- and non-match-to-sample. *Cognitive Psychology*, 86, 87−111.

Hochmann, J-R., Tuerk, A. S., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017). Children' s representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology*, 99, 17−43.

Johansen, M. K., Savage, J., Fouquet, N., & Shanks, D. R. (2015). Salience not status: How category labels influence feature inference. *Cognitive Science*, 39(7), 1594−1621.

Kalish, C. W., & Lawson, C. A. (2007). Negative evidence and inductive generalisation. *Thinking & Reasoning*, 13(4), 394−425.

Kemler-Nelson, D. G., Herron, L., & Holt, M. B. (2003). The sources of young children' s name innovations for novel artifacts. *Journal of Child Language*, 30(4), 823−843.

Kemler-Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development*, 71(5), 1271−1288.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, 83(1), 173−185.

Legare, C. H. (2014). The contribution of explanation and exploration to children' s scientific reasoning. *Child Development Perspectives*, 8(2), 101−106.

Li, F., Xie, L., Yang, X., & Cao, B. (2017). The effect of feedback and operational experience on children' s rule learning. *Frontiers in Psychology*, 8, 534.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44(2), 523−531.

Long, C. Q., Wu, R. M., Li, H., Chen, A. T., Feng, T. Y., & Li, F. H. (2006). 3.5~5.5 years older' s inductive reasoning in similarity vs conception confliction condition. *Acta Psychologica Sinica*, 38(1), 47−55. [龙长权, 吴睿明, 李红, 陈安涛, 冯廷勇, 李富洪. (2006). 3.5~5.5 岁儿童在知觉相似与概念冲突情形下的归纳推理. 心理学报, 38(1), 47−55.]

Nazzi, T., & Gopnik, A. (2000). A shift in children' s use of perceptual and causal cues to categorization. *Developmental Science*, 3(4), 389−396.

Nielsen, M., & Haun, D. (2016). Why developmental psychology is incomplete without comparative and cross-cultural perspectives. *Philosophical Transactions of the Royal Society B*, 371(1686), 20150071.

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.

Noles, N., & Gelman, S. (2012). Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology*, 48(3), 890–896.

Patnaik, A. (2008). *The physical discrimination and possible concept of object weight that exists in infants and toddlers* (Unpublished Master' s thesis). Retrieved from http://digitalcommons.wku.edu/theses/46/

Paulus, M., & Hauf, P. (2011). Infants' use of material properties to guide their actions with differently weighted objects. *Infant and Child Development*, 20(4), 423—436.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Povinelli, J. D. (2012). *World without weight: Perspective on an alien mind.* Oxford University Press.

Richland, L. E., Chan, T-K., Morrison, R. G., & Au, T. K-F. (2010). Young children's analogical reasoning across cultures: Similarities and differences. *Journal of Experimental Child Psychology*, 105(1—2), 146—153.

Schrauf, C., Call, J., & Pauen, S. (2011). The effect of plausible versus implausible balance scale feedback on the expectancies of 3- to 4-year-old children. *Journal of Cognition and Development*, 12(4), 518-536.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, 36(3), 273—310.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, 8(2), 159-182.

Walker, C., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductivity rich properties. *Cognition*, 133(2), 343-357.

Wang, X. Y., Yu, W. Y., Liang, P. P., Long, Z. J., Guo, X. L., Zhong, N., ···Zhou, H. Y. (2014). Influence of knowledge on label effect during adults' inductive reasoning. *Journal of Psychological Science*, 37(4), 845—850. [王旭艳, 于伟艳, 梁佩鹏, 龙周军, 郭晓龙, 钟宁, 秦裕林, 周海燕. (2014). 知识对成人归纳推理标签效应的影响. 心理科学, 37(4), 845—850.]

Wang, Z. (2014). *Children' s imitation of a weight sorting rule* (Unpublished Master' s thesis). Georgia State University. Retrieved from https://scholarworks.gsu.edu/psych_{theses}/126

Wang, Z., Meltzoff, A. N., & Williamson, R. A. (2015). Social learning promotes understanding of the physical world: Preschool children' s imitation of weight sorting. *Journal of Experimental Child Psychology*, 136, 82—91.

Wang, Z., Williamson, R. A., & Meltzoff, A. N. (2015). Imitation as a mechanism in cognitive development: A cross-cultural investigation of 4-year-old children' s rule learning. *Frontiers in Psychology*, 6:562.

Wang, Z., Williamson, R. A., & Meltzoff, A. N. (2018). Preschool physics: Using the invisible property of weight in causal reasoning tasks. *PloS ONE*, 13(3), e0192054.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology*, 36(5), 571—581.

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, 72(6), 1653–1673.

Wellman, H. M., & Gelman, S. A. (1988). Children' s understanding of the nonobvious. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*, Vol. 4, (pp. 99–135). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, 17(12), 1075–1081.

Note: Figure translations are in progress. See original paper for figures.