Title

Can unpredicted outcomes be intended? The role of outcome-beliefs in children's judgments of intention

Journal

Cognitive Development, 26(2)

ISSN

08852014

Authors

Liao, Yu
Li, Hong
Deak, Gedeon O

Publication Date

2011-04-01

DOI

10.1016/j.cogdev.2010.09.003

Peer reviewed

Cognitive Development 26 (2011) 106–117

Contents lists available at ScienceDirect

Cognitive Development

Can unpredicted outcomes be intended? The role of outcome-beliefs in children's judgments of intention

Yu Liao [a,b,c], Hong Li [a,b,*], Gedeon O. Deák [c]

[a] Key Laboratory of Cognition and Personality (SWU), Ministry of Education, Southwest University, Chongqing 400715, China
[b] School of Psychology, Southwest University, Chongqing 400715, China
[c] University of California-San Diego, Dept. of Cognitive Science, 9500 Gilman Dr., La Jolla, CA 92093-0515, United States

ARTICLE INFO

Keywords:
Causal inference
Positivity bias
Theory of Mind
Intention

ABSTRACT

An adult-like concept of intention includes a deliberate action to achieve a goal and a belief that one's action (if successful) will cause the desired outcome. For example, good outcomes caused by accident or by chance are not believed to be caused intentionally. In two experiments, we asked whether children understand this connection between intentions and outcomes. Children played two games in which actions could produce unintended outcomes (i.e., causes were unplanned). Children sometimes received a desirable reward independent of intention. In Experiment 1, 4- and 5-year-olds mistakenly claimed they had intended the desirable outcome even when it was unexpected. Four-year-olds judged that they had not intended a deliberate action if it did not yield a rewarding outcome. Experiment 2 demonstrates that 6-year-olds seldom make these errors. The results suggest that 4- and 5-year-old children have not yet attained an adult-like concept of intention. Their inaccurate judgments regarding their intentions, given a rewarding yet unexpected outcome, can be explained by a positivity bias.

Humans often evaluate whether the results of their actions fit prior intentions. This evaluation is crucial for judging performance, learning from errors, and updating plans and goals. We also frequently infer the intentions of other people's actions and judge whether their results were intended. Understanding intention is a fundamental aspect of social cognition, critical to reasoning about our own and other people's actions, past, present, and future (Moses, 1993). To understand intentions we

* Corresponding author at: Key Laboratory of Cognition and Personality of Ministry of Education, Southwest University, No. 1 Tian Sheng Rd., Beibei, Chongqing 400715, China. Tel.: +86 23 68253625.
E-mail addresses: lihong1@swu.edu.cn, lihongwrm@vip.sina.com (H. Li), deak@cogsci.ucsd.edu (G.O. Deák).

0885-2014/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogdev.2010.09.003

must coordinate representations of desires, beliefs, actions, and outcomes (Malle & Knobe, 1997). An intention connects a goal to fulfill some desire with a belief that some planned action will achieve the desired outcome. We judge an intention to be fulfilled only if the outcome was caused by the action expected to fulfill it. An understanding of intention is called for when the planned action is fully expected (believed) to be completed, yet fails to satisfy the desire. An even more nuanced understanding is required when a planned action does not yield the intended outcome, but a desirable outcome results nonetheless.

For example, if you rummage for a pencil you are sure is on your desk and fail to find it, this violates your belief that your intended action will cause the desired outcome. If your search unexpectedly reveals a lost gift certificate, the outcome is certainly desirable, but not the one you had intended. If you give up looking for the pencil, but then step on it unexpectedly, you have attained the desired outcome, but *not* due to your intentional actions. We cannot say that we intended to find the pencil by stepping on it. Similarly, walking away was not an action *intended* to fulfill the desire for a pencil. Recognizing this distinction requires understanding that an intended outcome is achieved only if it occurs in an expected way. Thus, we could not have intended a result that we did not know was possible. To make sense of such situations, we must distinguish between *intended* outcomes and the *actual* outcomes caused by our motivated actions.

Children's understanding of intention undergoes a protracted development. Young children appear to struggle with understanding mismatches (Astington, 2001). For example, 3-year-olds can correctly state an actor's intentions (Astington, 1993; Lang & Perner, 2002; Liao et al., 2006; Shultz & Wells, 1985), but only when the result of an intended action matches the desirability of that outcome (Astington, 2001). For example, if a young child at a carnival tosses a ring on the peg she is aiming for and wins a prize that she desires, she could report that her action was intentional. However, if she successfully hits the peg but it turns out not to be a winning peg, she might say that she had not been trying to hit that one. Doing so reveals her difficulty resolving a mismatch between intention and desired outcome. Conversely, if she misses the intended peg but hits a more valuable one (and receives a larger-than-expected prize), she might mistakenly say that she had been trying to hit the more valuable peg. This also demonstrates difficulty resolving a mismatch. In both cases an adult would recognize that the original intention did not match the outcome, whether undesirable (no prize) or desirable (bigger prize).

Children aged 3–4 seem to handle mismatches of intentions and outcomes by making desire-based judgments. Schult (2002) asked 3-, 4-, and 5-year-olds to throw beanbags at colored buckets, one of which contained a small picture that served as a reward. Children had to say which bucket they intended to hit before tossing. If the bucket they hit contained a picture, they received a prize. Because children did not always hit the intended bucket, some trials yielded mismatches. For example, children could hit the "wrong" (i.e., unintended) bucket but still get a reward. In these trials, the intention was unfulfilled but the outcome desirable. Alternatively, they could hit the right bucket but get no reward, so that the intention was fulfilled but the outcome undesirable. Both of these trials reflect mismatches between fulfillment of intentions and desirability of outcomes. After each trial, children were asked which bucket they had tried to hit. All age groups answered correctly when intentions and outcome matched (i.e., fulfilled-and-desirable or unfulfilled-and-undesirable). Thus, children could at a minimum correctly represent the desired outcome. However, in mismatch trials where children hit an unintended bucket but still received a reward, 3-year-olds and some 4-year-olds claimed that they had intended to hit the rewarding bucket. That is, they misreported their original intention if the outcome was unexpectedly desirable. This suggests a desire-based concept of intention. In a related study by Phillips, Baron-Cohen, and Rutter (1998), 4- and 5-year-olds fired "ray-guns" at cans that were rigged so that an experimenter controlled which can fell and whether or not it contained a prize. As in the Schult study (2002), four conditions represented matches vs. mismatches and desirable vs. undesirable outcomes. Five-year-olds could report their intentions even when the outcome was mismatched. However, some 4-year-olds made many errors, especially when they hit the intended can but received no reward. Thus, children had difficulty reconciling mismatches between intentions and unexpected outcomes.

These results suggest that 4- and 5-year-olds only gradually learn to recognize and reconcile mismatches between fulfillment of intentions and desirability of outcomes (Feinfield, Lee, Flavell, Green,

& Flavell, 1999). However, it is not clear what conceptual insight(s) children acquire and at what age. One possibility is that younger children simply report the desirability of the outcome and "rewrite history" to conform their original intention to the outcome's desirability. However, Schult's results (2002) cannot be explained solely by desire-based representations. Children had more difficulty in mismatch trials where an intention was unfulfilled but they were rewarded than in trials where they hit the intended bucket but got no reward. If one type of mismatch was more difficult, some factor other than desire must have contributed.[1]

One potential factor is children's *positivity bias* concerning the efficacy of their actions. Specifically, children tend to attribute good outcomes more than bad outcomes to an actor's own efforts or skills (Kunda, 1987; Mezulis, Abramson, Hyde, & Hankin, 2004; Whitley & Frieze, 1985). Thus, when the cause of a desirable outcome is ambiguous, children might assume that their deliberate actions were responsible. This could explain the asymmetry found by Schult. This bias goes beyond simple desire-based responding because it incorporates beliefs about the causal relation between planned actions and outcomes. These beliefs might make 3- and 4-year-olds especially likely to rewrite the history of their intentions when they experience a desirable outcome.

Adults also have a self-serving positivity bias (Mezulis et al., 2004), but they do not make mismatch errors. Apparently, the "mature" concept of intention (Searle, 1983) stipulates that intentions are limited by one's beliefs about the possible results of one's planned actions. In the Schult study (2002), very few 5-year-olds made errors answering the test questions. However, Schult's paradigm introduces a factor that complicates interpretation of his results: the task required skill. Children sometimes missed their targets during training and test trials. This would have made them uncertain about the outcome of their actions, because either the intended or unintended outcome was possible on any trial. This uncertainty about the likelihood of completing the intended action might lead children to maintain a less coherent representation of their actions and therefore a less coherent representation of the mismatch between actions and outcomes.

Because we know neither the magnitude of this represented mismatch, nor to what degree positivity bias influenced children's beliefs in this skill-based task, we also do not know how belief disconfirmation affected children's answers in each mismatch condition. A purer test would be to examine whether children know that they could not have intended an outcome they did not believe possible. This removes any effect of mixed expectations about possible outcomes.

The skill requirement of Schult's (2002) bucket task might have contributed more subtly to the difficulty of unfulfilled/desirable mismatches. Children's desire might be partially satisfied simply by hitting one of the buckets, given that this is a challenging task. During training, children may have experienced failed attempts not only to hit the intended bucket, but also to hit *any* bucket at all. In a previous, unpublished replication, we observed that children sometimes failed to hit any bucket. Thus, hitting any bucket became a secondary desire. This desire was fulfilled in "unfulfilled/desirable" mismatch trials, so children might have reported (correctly) that they tried to at least hit *some* bucket, even if they missed the exact one they aimed for. However, these answers were counted as errors. No such error would be made in fulfilled/unrewarded mismatch trials. Thus, a secondary desire to hit any bucket might have contributed to the asymmetry between the two mismatch types. It is therefore difficult to test the positivity bias account using skill-based tasks like those used by Schult (2002) or Phillips et al. (1998).

For these reasons, we sought to test children's difficulty representing intentions and mismatched outcomes to determine whether, and why, children have greater difficulty making the correct judgments when unintended outcomes are rewarded. In our paradigm, children played two games in which their actions could yield completely unexpected and rewarded outcomes. Critically, each game was simple and did not require any skill. Thus, unintended outcomes were truly completely unexpected:

[1] Phillips also reported an asymmetry in the difficulty of the two mismatch trials. However, they found that 4-year-olds had more trouble in mismatch trials where an intention was fulfilled but they were not rewarded. Although this is the opposite asymmetry, as the authors noted, in their task children might have said they meant to hit the other color hoping that the experimenter would simply give them the prize. Also, their question was "Did you mean to hit ... or ...?" This less common sense of "mean" could be unfamiliar to young children and could have been interpreted as referring to their desire rather than their intention.

they violated children's beliefs about the possible outcomes of their intended action. To ensure that children would never come to believe that an unfulfilled/rewarded outcome was possible, we presented only one unintended outcome per game. By using two games, we were able to meet this constraint while also ensuring that children's responses generalize beyond one specific situation.

In both games, the experimenter secretly controlled the matching or mismatching of intentions and desired outcomes on each trial. In the "secret passage" game, children were shown a box with two holes on the top and two doors on the bottom. A prize was hidden behind one door. Children learned that if they put a ball in one of the top holes, it would fall and open the door directly below it. However, unbeknownst to them, in some trials a secret diagonal passage was activated to change the path of the ball to the opposite door, making that door open unexpectedly. Children were trained to choose the hole that they believed was above the door that hid the prize. Dropping the ball required no skill. However, by covertly manipulating the ball's direction and the location of the prize, the experimenter could control whether the child's intention was fulfilled and whether the outcome was rewarded. Children were then asked to state which door they had intended to open. Since children did not know about the secret passage, they could not have believed it possible for the ball to open the opposite door. These trials assessed children's understanding that they could not have intended an outcome they never believed possible, even if the outcome was desirable.

The second, "secret magnet" game was analogous. Children lowered a toy puppy into one of two rooms in order to find a hidden reward, a "pig" sticker. The experimenter could, however, use a hidden magnet to guide the puppy into the unintended room as the child lowered it. The experimenter also manipulated whether or not the child found the sticker in the unintended room. This design allowed any match or mismatch of intention and desirability.

The critical trials in each game combined unintended results with desirable outcomes. To answer correctly, children needed to inhibit any positivity bias (i.e., the expectation that their actions would cause positive results) and refer to their own prior beliefs about what outcomes were possible. They needed to realize, in effect, "I did not intend a result that I did not know was possible." The opposite mismatch, where children's intention is fulfilled (e.g., the ball falls directly downward) but the reward sticker is not in the expected location, was clearly possible: "I hit it, but the reward was hidden in the other place." Because the positivity bias should not strongly relate to random or fortuitous rewards, which are less causally related to one's own agency, these mismatches might be easier for young children to resolve. Thus, if young children make more errors when unfulfilled intentions are followed by desirable results, it would imply that children come to understand that their intentions can fail (or fail to yield a pay-out) before they realize that one cannot have intended an unimagined outcome. By contrast, if children make many errors in response to both kinds of mismatches, it would simply imply a desire-based concept of intention with no added role of biased beliefs about causality. We hypothesize that children progress from having a desire-based concept to incorporating beliefs about causality skewed by the positivity bias, to forming a concept of intention that entails self as a causal agent. If so, the order of difficulty for children should be matching trials (least difficult), then intended/undesirable mismatch trials, and finally unintended/desirable trials.

1. Experiment 1

1.1. Method

1.1.1. Participants

Thirty-two 4-year-olds (14 females; mean age 4–6, range 4-1 to 4-11) and 33 5-year-olds (17 females; mean age 5–3, range 5-1 to 5-10) were recruited from a kindergarten affiliated with Southwest University in Beibei, Chongqing, China. Most parents were university employees. All children were Han ethnic Chinese and fluent speakers of standard Mandarin, and none was reported to have a cognitive, sensory, or language deficit. Children were tested individually in a quiet room at their school.

1.1.2. Design

By manipulating the outcomes of children's intentions (fulfilled/unfulfilled) and whether or not they received a desirable reward, we created four conditions for each task: (1) intention fulfilled and

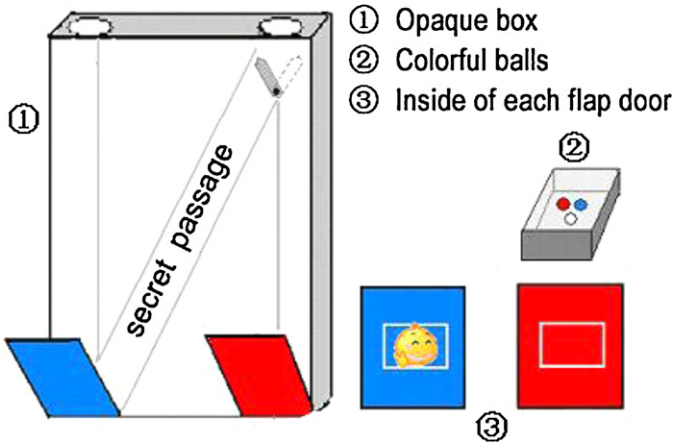

Fig. 1. Apparatus for secret passage game. (1) Opaque box, shown in cross-section (i.e., internal tubes were not visible), with holes and doors with flaps. (2) Several small colorful balls. (3) Transparent pocket to enclose prize sticker.

reward received (I+R+); (2) intention fulfilled but no reward (I+R−); (3) intention unfulfilled but reward received (I−R+); (4) intention unfulfilled and no reward (I−R−). Conditions 2 and 3 each presented a mismatch between intention-fulfillment and outcome-desirability.

Each child completed two games. In each game, children experienced only one unexpected outcome, and therefore never came to expect an "impossible" outcome. Participants in each age group were randomly assigned to one of two between-subject conditions, matching or mismatched. One received I+R+ and I−R− trials (i.e., matching), with one trial of each type per task. The other received I+R− and I−R+ trials (i.e., mismatched), also with one trial of each type per task. The design not only limited unexpected outcomes to one per child; it also showed whether children responded using a desire-based concept yielding all correct answers in the first group (I+R+, I−R−) and all errors in the second group (I+R−, I−R+).

Each child completed at least four practice trials per game with no test questions. The practice trials were all I+ but outcomes were a mix of R+ and R−.

In test trials, I+ trials always came before I− trials (i.e., the magnet or diagonal passage would not be used until after at least one I+ trial). Test questions were asked after each trial. Children's scores were the sum of their correct answers in both games, ranging from 0 to 2 per condition per game.

1.1.3. Materials: secret passage game

As shown in Fig. 1, the secret passage game included an opaque box and several small colorful balls. Three tubes inside the box connected the holes and doors. Two were vertical, connecting each hole to the exit directly below. The third, "secret" passage was diagonal, connecting the right hole to the left door (or, for half of the children, the left hole to the right door, to control for side preferences). A secret internal gate operated by the experimenter controlled whether the vertical or diagonal passage was open. The doors were covered by a blue and a red flap (side counterbalanced) and were closed until one was pushed opened by the ball's exit. Each door had a transparent pocket inside. A prize placed in one pocket (a sticker) became visible when the door opened.

1.1.4. Materials: secret magnet game

The apparatus included a wooden box with transparent windows on the front and sides and two holes on top that opened into two small compartments – described as rooms for the puppy – below. The floors of the rooms were covered with blue and red paper (side counterbalanced). A blank sticker and a sticker with a picture of a pig were placed in each room face down. A toy puppy with a magnet inside was attached to a string; another magnet was hidden in a compartment under one of the rooms during "unexpected" trials.

1.1.5. Procedure: secret passage game

The experimenter first showed children how to play the game, then gave children a ball and asked, "Can you use this ball to open the red/blue door?" Children tried opening each door at least once. The secret passage was not used. This activity was designed to teach children that balls would fall vertically to the door below.

Children were then introduced to a "finding game." They were shown an attractive sticker and told that it would be hidden behind one door, and if that door opened they could keep it. The experimenter then asked, "Which door will you try to open this time?" and hid the sticker in the appropriate place. After children answered, they dropped the ball into the specified hole. Each trial conformed to one of four conditions, based on whether the secret passage was activated or not and on whether the sticker was behind the door chosen. Three test questions were asked after each trial, in counterbalanced order. Children received one point per correct answer. All instructions and questions were read and critiqued by six Chinese developmental psychologists and by the two Mandarin-fluent authors, and wording was determined to be age-appropriate. (For a complete transcript in Chinese of all instructions and questions, contact the authors.)

Intention question: "Which door did you try to get the ball to open: the [red/blue] one or the [blue/red] one?" (Color-word order was counterbalanced in all questions.)

Desire question: "Which door do you really wish the ball had opened: the [red/blue] one or the [blue/red] one?"

Prior Belief question: "When you put the ball in this hole [pointing to the chosen hole], which door did you think would open: the [red/blue] one or the [blue/red] one?"

1.1.6. Procedure: secret magnet game

The experimenter first introduced "Piggy's big room" and showed children how to lower the puppy by a string from the left or right hole into the room below. The experimenter then trained children with the apparatus by asking, "Can you send the puppy to the [red/blue] room?" Children sent the puppy to each room at least once. The magnet was not used so that, after this training activity, children would expect the puppy to descend into the room directly below each hole.

Children were then taught a "find Piggy" game. A blank sticker and the Piggy sticker were each placed face down in one of the two rooms so that children could not see which room contained which sticker. In each trial, the children chose a room. If the puppy found Piggy, the child received the sticker as a prize. Children were given the puppy and asked, "Which room will you send puppy to?" Children then lowered the puppy. If the magnet was placed under the non-chosen room, the puppy unexpectedly swerved to that room (I− condition). If the Piggy sticker was revealed, the child got the reward (R+). The same three test questions were asked after each trial, in counterbalanced order, with color-word order counterbalanced:

Intention question: "Which room did you try to send the puppy to: the [red/blue] one or the [blue/red] one?"

Desire question: "Which room do you really wish the puppy had gone to: the [red/blue] one or the [blue/red] one?"

Prior Belief question: "When you put the puppy in this door [experimenter pointed], which room did you think the puppy would be sent to: the [red/blue] one or the [blue/red] one?"

1.2. Results

Analyses were conducted for each question type, using a 2 (intention: fulfilled or unfulfilled) × 2 (matching or non-matching) × 3 (test sequence: first, second or third question) mixed-measures analysis of variance (ANOVA). No main effect for test sequence was significant: the F values for this factor for the intention, desire, and prior belief questions were, respectively, $F < 1$, $F < 1$, and $F = 2.87$ ($p = .065$; $\eta 2 = .089$). No interaction between test sequence and any other factor appeared. Thus, question sequence did not affect performance.

Within-subject *t*-tests showed that children's accuracy did not differ between games in any condition, indicating that the games are comparable. Thus, results from the two games are collapsed.

Table 1
Mean number of correct answers for test questions, by age, Experiment 1.

Questions	Age	I+R+	I+R−	I−R+	I−R−
Try	4-year-olds	1.73 (.59)	1.00 (.79)	.35 (.49)	1.53 (.64)
	5-year-olds	1.88 (.33)	1.31 (.70)	1.00 (.81)	1.65 (.61)
Desire	4-year-olds	1.53 (.74)	1.59 (.62)	1.88 (.33)	1.67 (.72)
	5-year-olds	1.76 (.44)	1.81 (.54)	1.69 (.60)	1.71 (.59)
P-Belief	4-year-olds	1.87 (.52)	1.00 (.87)	.35 (.60)	1.47 (.74)
	5-year-olds	1.88 (.49)	1.31 (.71)	.37 (.62)	1.29 (.85)

Note: Values in parentheses are standard deviations.

Children's accuracy by question and condition is shown in Table 1. A 3 (question type: desire, intend, or believe) × 2 (matching: match [I+R+ and I−R−] or mismatch [I+R− and I−R+]) × 2 (intention: fulfilled [I+] or unfulfilled [I−]) mixed-model ANOVA was conducted. Question type and intention were within-subjects conditions; matching was between subjects.

1.2.1. Overall condition and question effects

All main effects were significant. Children performed differently across test questions, $F(2,62) = 28.82$, $p < .001$, $\eta 2 = .31$. A post hoc LSD test showed that children answered desire questions more accurately than both intention questions ($p = .001$) and prior-belief questions ($p = .001$). No differences were found between intention and prior-belief questions, suggesting that errors were not due to use of specific mental verbs, but rather to children's representations of their mental state when selecting their action.

The main effect of intention-fulfillment was significant, $F(2,62) = 29.37$, $p < .001$, $\eta 2 = .32$: children were more accurate when the intention they expected to enact was fulfilled. This confirms that children had difficulty reconciling unexpected outcomes with prior beliefs. Matching vs. mismatching was also a significant factor, $F(2,62) = 34.90$, $p < .001$, $\eta 2 = .36$: children were more accurate when intention-fulfillment matched the desirability of the outcome (i.e., reward).

There was a question-by-matching interaction $F(2,126) = 26.73$, $p < .001$ $\eta 2 = .30$. The mismatch group performed differently across questions, $F(2,64) = 47.78$, $p < .001$, $\eta 2 = .60$. Follow-up tests showed that they were more accurate when answering desire questions than intention or prior-belief questions (both $p < .001$). In contrast, there were no differences between questions in the matching group.

There was also a significant question-by-intention-fulfillment interaction $F(2,126) = 11.63$, $p < .001$, $\eta 2 = .156$. A post hoc LSD test showed that on unfulfilled trials, intention and prior belief questions were more difficult than desire questions ($p < .001$; $p = .022$, respectively). However, when intentions were fulfilled, children performed equally well on all three questions.

Overall, children performed better on the desire questions than on the other questions. Intention and prior-belief questions were more difficult when intentions were unfulfilled or when there were mismatching outcomes.

1.2.2. Age effects

A 2 (age: 4 vs. 5) × 2 (match vs. mismatch) × 2 (intention fulfilled vs. unfulfilled) mixed-model ANOVA was conducted for each test question.

For the desire question, results showed no main effects or interactions. Both age groups were generally accurate in all conditions. Correct answers in I+R+, I+R−, I−R+, and I−R− trials averaged 71.9%, 75.8%, 81.8%, and 78.1%, respectively. This confirms that children wanted the sticker and accurately registered whether or not they received it.

All main effects were significant for reports of prior intention. As shown in Fig. 2, overall, 5-year-olds did better than 4-year-olds, $F(1,61) = 5.91$, $p < .05$, $\eta 2 = .088$. Children answered more accurately when intention and desire matched, $F(1,61) = 38.76$, $p < .001$, $\eta 2 = .389$. They also answered more accurately when the intention was fulfilled, $F(1,61) = 13.07$, $p = .001$, $\eta 2 = .177$. There were no significant interactions. In paired comparisons, questions ranked from least to most difficult as follows: I+R+, I−R−, I+R−, I−R+. All paired comparisons were significant, except I+R+ and I−R−; the two matching conditions

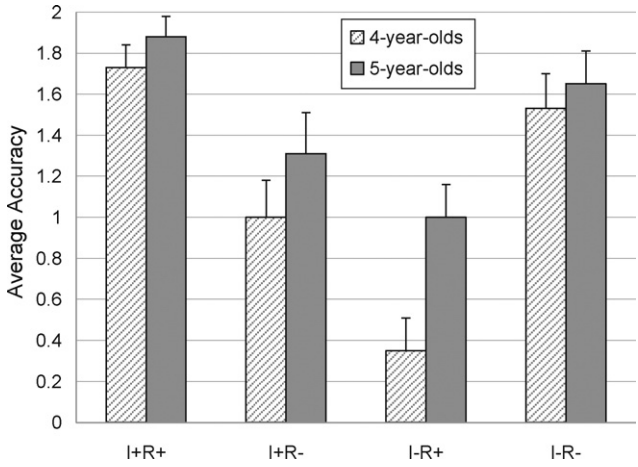

Fig. 2. Children's mean accuracy on the intention question in Experiment 1, by age and condition (bars = standard error).

were only marginally different ($p = .051$). Both 4-year-olds and 5-year-olds were more accurate in the I+R+ and I−R− conditions than expected by chance (expected mean = 1.0). As predicted, the critical finding was a difference between the two mismatch conditions, $F(1,31) = 9.29$, $p < .01$, $\eta 2 = .231$: it was more difficult for children to report their intention when it was unfulfilled but still followed by a reward. In this condition, 4-year-olds' performance was below chance, and 5-year-olds' was not different from chance. Age groups were compared in each condition using post hoc tests. Five-year-olds performed better than 4-year-olds in I−R+ trials $F(1,31) = 7.706$, $p < .01$, $\eta 2 = .199$. No other significant differences were found.

For the prior-belief question, there were main effects of matching and intention-fulfillment, but no interactions. Children were more accurate when their intentions were fulfilled $F(1,61) = 28.94$, $p < .001$, $\eta 2 = .322$, and when the fulfillment of their intentions matched the desirability of outcomes, $F(1,61) = 50.32$, $p < .001$, $\eta 2 = .45$. Questions across conditions ranked, from least to most difficult, as follows: I+R+, I−R−, I+R−, and I−R+. All pairs were significantly different.

1.2.3. Correlations between test questions

If children's prior beliefs constrain their post hoc judgments about their intentions, there should be a correlation between them. In addition, if it is not fully reducible to a desire-based effect, there should be a smaller or negligible correlation with that question. Because prior belief was true in the I+ conditions and false in the I− conditions, we calculated the correlations between test questions separately in the two conditions. Significant correlations were found between intention and prior-belief questions, both when the intention was fulfilled ($r = .505$, $p < .001$) and unfulfilled ($r = .444$, $p < .001$). No other significant correlations were found.

1.3. Discussion

Children's understanding of intention was assessed by varying whether the intention was satisfied and whether a desirable reward was received. In the critical conditions, the outcomes did not match intentions; I−R+ trials featured a surprising, unintended but rewarding outcome. Children were asked to report what they had wanted, what they had intended to do, and what they had believed would happen.

As predicted, it was not difficult for 4- and 5-year-olds to report their intentions when the success of the intention matched the desirability of the results (i.e., I+R+ and I−R−). In contrast, when success and desirability are mismatched (i.e., I+R− and I−R+), reference to desire causes errors. In I−R+ trials, children must recognize that the desirable outcome was not caused by their planned action; rather, it came about in an unexpected way. As predicted, these trials were more difficult than I+R− mismatches,

suggesting that it is not as difficult to represent in the I+R− case that the original intention was carried out, even if no reward was received. This is, of course, a common occurrence in everyday life, and even toddlers show different emotional reactions to the successes or failures of their actions to achieve desired outcomes, even if no extrinsic rewards are at stake (Bullock & Lütkenhaus, 1988).

Accuracy on prior-belief questions was highly correlated with intention accuracy whether the intention was fulfilled or unfulfilled. This shows that errors were truly related to children's beliefs about how their own deliberate actions would cause some result. Children only knew that their prior belief was confirmed when the desired outcome was obtained. Thus, many children reported that they had believed, for example, that the puppy would swerve into the red room (something they did not know was possible) if they found the desired piggy sticker there. This suggests that children's difficulties in representing their own false beliefs are related to their difficulties in representing their own past intentions.

All participants answered the desire question accurately, showing that they had wanted the sticker (i.e., it was an effective reward), and they remembered this desire. This suggests that the results cannot be due to children rewriting the history of what they wanted (e.g., "Hmph, I didn't want that dumb sticker anyway!").

In two mismatch trials even 5-year-olds did not perform above chance, and on I−R+ prior-belief questions they were less accurate than expected by chance. This contrasts with Schult's (2002) finding, in which 5-year-olds performed better than chance. Thus, when children have no reason to expect their intention to fail (because the task does not require skill, and they have not seen it fail in a surprising way), testing reveals residual limitations in older children's concept of intention. Perhaps when their easy action fails they cannot make sense of this surprising outcome, and when they are rewarded nonetheless, they resort to the positivity bias. That is, failing to understand how the ball or dog ended up in a different place, 5-year-olds falsely assume that they were actually trying to produce the rewarded result.

Before accepting this conclusion, we should rule out an alternative explanation. Perhaps this task is very difficult, even for older children, because the unexpected result is so surprising. Although the results from the desire question and from the I−R− trials seem to disconfirm this, it is still possible that some aspect of the games or the questions makes them difficult even for older children. To rule out this possibility, in Experiment 2 we tested 6-year-olds, who should also be surprised by the results in the I− trials, but who have been shown to have no difficulty reconciling intention-outcome mismatches.

2. Experiment 2

We replicated Experiment 1, this time extending the age range to include 6-year-olds and making several changes to eliminate extraneous sources of difficulty. Most notably, there is evidence that asking young children several questions about the same topic can encourage perseverative responses (Deák, 2006). Thus, we asked only the critical "try" question. Also, to limit task duration and ensure children's continued attention, we dropped I+R+ trials. Recall that even 4-year-olds were near ceiling on these trials.

If 6-year-olds perform well on I−R+ trials, it would suggest that young elementary school-aged children have attained a concept of intention that is, in an important regard, similar to that of adults.

2.1. Method

2.1.1. Participants

Twenty 4-year-olds (12 females; mean age 4–6, range 4-1 to 4-11); 19 5-year-olds (eight females; mean age 5–5, range 5-0 to 5-11) and 22 6-year-olds (12 females; mean age 6–6, range 6-1 to 6-11) were recruited from the same population and in the same way as in Experiment 1. None had participated in Experiment 1.

2.1.2. Experimental design and procedure

The secret passage and secret magnet games were administered as in Experiment 1, but only the intention question was asked, and only three conditions (I+R−, I−R+, I−R−) were administered.

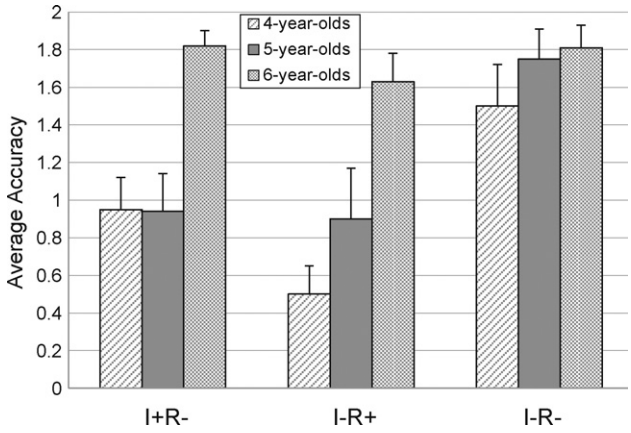

Fig. 3. Children's mean accuracy on the intention question in Experiment 2, by age and condition (bars = standard error).

Participants in each age group were randomly assigned to two groups. All participants completed an I+R− trial. Group 1 then completed the other mismatch trial, I−R+ (unfulfilled/rewarded). Group 2 completed the other unfulfilled intention, no reward (I−R−) trial. Again, each child saw only one unexpected outcome.

2.2. Results and discussion

Children's performance in each condition is shown in Fig. 3. Because all children completed an I+R− trial, data for this trial type include all participants. The average for I−R+ trials is from Group 1 children, and the average for I−R− is from Group 2 children. A 2 (group) × 3 (age) ANOVA on I+R− trials showed no group difference, and no significant group × age interactions.

Age differences were significant in I+R− trials, $F(2,58) = 8.50$, $p = .001$, $\eta 2 = .23$. Most 6-year-olds were accurate: 16 of 22 (72.7%) correctly reported their intention in both games; the rest were correct in one game. Post hoc LSD tests showed that 6-year-olds were more accurate ($M = 1.73$, SD = .45) than 5-year-olds ($M = .94$, SD = .85; $p = .020$) and 4-year-olds ($M = .90$, SD = .85; $p < .001$). Four- and 5-year-olds did not differ and were not above chance.

Age differences in the critical I−R+ trials were also significant $F(2,29) = 8.99$, $p < .001$, $\eta 2 = .38$. Most 6-year-olds (8 of 11, 72.7%) correctly responded in both games, 2 (18.1%) were correct in one game, and 1 (9.0%) failed both games. Post Hoc LSD tests showed that 6-year-olds were more accurate ($M = 1.63$, SD = .67) than 5-year-olds ($M = .90$, SD = .87; $p = .033$), and 4-year-olds ($M = .50$, SD = .52; $p < .001$), and 5-year-olds were more accurate than 4-year-olds, $p = .049$. Five-year-olds were not above chance and 4-year-olds were lower than chance, $p = .005$.

Age differences were not significant in I−R− trials. Average scores of 4-, 5-, and 6-year-olds were 1.5 (SD = .70), 1.75 (.46), and 1.81 (.40), respectively. Most children (21 of 29, or 72.4%) passed both games, 7 (24.1%) passed one game, and 1 (3.4%) failed both games. Children in each age group performed above chance.

Overall, the relative difficulty across trial types for 4- and 5-year-olds was the same as in Experiment 1. The easiest condition was I−R− ($M = 80\%$ correct), where intention was unfulfilled and desire was unattained. As predicted, children performed less accurately in I+R− trials ($M = 41\%$), $F(1,17) = 8.77$, $p < .01$, $\eta 2 = .34$, and least accurately in I−R+ trials ($M = 35\%$), $F(1,37) = 17.41$, $p < .001$, $\eta 2 = .32$. However, these two proportions did not differ significantly.

3. General discussion

These results extend previous findings that young children do not have a fully mature concept of intention. Previous studies found that 3-year-olds can report their intentions (Astington, 1993;

Joseph & Tager-Flusberg, 1999; Shultz & Wells, 1985), but only if they have received the reward they wanted (Schult, 2002). We investigated a more advanced understanding that involves reference to one's causal beliefs. We specifically asked when children know that intention *requires* a prior belief that one's deliberate actions will cause the desired outcome. At what age do children understand they could not have *intended* an outcome that they did not foresee or even know was possible? Four- and 5-year-olds in two games reported that they had intended an unimagined outcome (e.g., dog swerving into the blue room) that was qualitatively different than the outcome that they had planned (e.g., lowering the dog into the red room). This was not simply because they generally believe that all outcomes were intended: the error was more likely when a reward was received, even if they had learned that the reward was unpredictable. They made fewer errors (yet were not completely accurate) if the unintended outcome carried no reward. Thus, children had difficulty reporting an unintended result, particularly if something good came of it.

We therefore confirmed Schult's (2002) finding of asymmetry in the difficulty of mismatched intentions and outcomes: I−R+ trials were more difficult than I+R− trials. This asymmetry shows that the difficulty is not due merely to a desire-based concept of intention, because that would have yielded errors on all mismatch trials. However, the reason for this pattern of errors in Schult's data is ambiguous, because children knew that they might miss the intended bucket – or indeed any bucket. Thus, even hitting the wrong bucket might have satisfied a secondary desire not to miss completely. This makes "wrong" answers to I−R+ questions difficult to interpret. In our task, children had no idea that their intentions could be unfulfilled. Thus, answering I−R+ questions correctly required tracking one's original intention after receiving the desired reward, even when the *manner* in which the action was rewarded was fully unplanned and unexpected.

This explanation involves children's beliefs about the causal relation of their intentions to actions and outcomes. Children, like adults, are predisposed to believe that positive outcomes with ambiguous causes occurred by virtue of their own abilities or efforts. This positivity bias would be activated by a rewarding outcome, when the cause of that outcome is particularly ambiguous or hard to understand. This claim – that a developing concept of intention involves the refinement of beliefs about the causes and effect of people's actions – is supported by the parallel pattern of errors on prior-belief questions. These errors confirm that 4- and 5-year-olds' evaluations altered post hoc their beliefs about possible outcomes.

Experiment 2 shows that these challenges are not due to the intrinsic difficulty of the task. Six-year-olds performed accurately in every trial type and performed significantly better in the mismatch trials than 4- and 5-year-olds. By age 6, then, most children understand that they can only intend actions that they believe might possibly succeed.

The results are consistent with other findings on the development of mental-state inferences. Most notably, Taylor, Esbensen, and Bennett (1994) found that after 4- and 5-year-olds learned some novel, esoteric fact, they claimed to have known it for a long time. Similarly, children have difficulty reporting their own prior false beliefs (Astington, 1993). This sort of "rewriting" is similar to our participants' tendency to state that they intended to produce an unexpected outcome if the result was positive.

Several questions remain. First, are there I−R+ situations in which 5-year-olds might more easily report their intentions correctly? Notably, a minority of 5-year-olds in this study did answer I−R+ questions correctly. It would be useful to study these children's cognitive processes (e.g., memory for actions and results) to understand individual differences. Second, what situational variables might lead older children or adults to make the same errors as 4- and 5-year-olds? These could be not only situational variables (e.g., elapsed time), but also personality traits like defensiveness. A final question concerns cultural factors. This study tested urban, middle-class Chinese children. It is unknown whether the development of the concept of intention is language- or culture-bound. Western, English-speaking children produce results similar to ours: they do not differentiate intention from desire until 4.5 or 5 years of age (Feinfield et al., 1999; Phillips et al., 1998; Schult, 2002). Nonetheless, possible differences between Chinese- and English-speaking children require future cross-cultural studies In any case, our results show that the concept of intention continues to develop through the preschool years.

Acknowledgements

This work was supported by the Key Discipline Fund of National 211 Project, China (NSKD08014 to H. Li), and by the National Science Foundation's Human Social Dynamics program (SES-0527756 to G. Deák). Thanks to Zhenlin Wang and Ting Zhang for helpful discussions, Qin Zhang for assistance in collecting data, and Marybel Robledo for assistance in manuscript preparation. Many thanks to the principals, teachers, and students of the Southwest University kindergarten.

References

Astington, J. W. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard Univ. Press.

Astington, J. W. (2001). The paradox of intention: Assessing children's meta-representational understanding. In B. Malle, L. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 85–103). Cambridge, MA: MIT Press.

Bullock, M., & Lütkenhaus, P. (1988). The development of volitional behavior in the toddler years. *Child Development*, *59*, 664–674.

Deák, G. O. (2006). Do children really confuse appearance and reality? *Trends in Cognitive Sciences*, *10*, 546–550.

Feinfield, K. A., Lee, P. P., Flavell, E. R., Green, F. L., & Flavell, J. H. (1999). Young children's understanding of intention. *Cognitive Development*, *14*, 463–486.

Joseph, R. M., & Tager-Flusberg, H. (1999). Preschool children's understanding of the desire and knowledge constraints on intended action. *British Journal of Developmental Psychology*, *17*, 221–243.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*, 636–647.

Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology*, *20*, 67–76.

Liao, Y., Wu, R., Zelazo, P. D., Li, H., Zhang, T., Zhang, L., et al. (2006). Relations among test questions in unexpected transfer task, understanding of intention and executive function. *Acta Psychologica Sinica*, *38*(2), 207–215 [in Chinese].

Malle, B. F., & Knobe, J. (1997). The folk concept intentionality. *Journal of Experimental Social Psychology*, *33*, 101–121.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*, 711–747.

Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development*, *8*, 1–25.

Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology*, *16*, 337–348.

Schult, C. A. (2002). Children's understanding of the distinction between intention and desire. *Child Development*, *73*, 1727–1747.

Searle, J. R. (1983). *Intentionality: An essay in philosophy of mind*. Cambridge Univ. Press.

Shultz, T. R., & Wells, D. (1985). Judging the intentionality of action-outcomes. *Developmental Psychology*, *21*, 83–89.

Taylor, M., Esbenson, B. M., & Bennett, R. T. (1994). Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development*, *65*, 1581–1604.

Whitley, B. E., & Frieze, I. H. (1985). Children's causal attributions for success and failure in achievement settings: A meta-analysis. *Journal of Educational Psychology*, *77*, 608–616.