Can’t stop believing: inhibitory control and resistance to misleading testimony

Vikram K. Jaswal, 1 Koraly Pérez-Edgar, 2 Robyn L. Kondrad, 3 Carolyn M. Palmquist, 4 Caitlin A. Cole 5 and Claire E. Cole 6

1. Department of Psychology, University of Virginia, USA
2. Department of Psychology, The Pennsylvania State University, USA
3. Department of Psychology, Appalachian State University, USA
4. Department of Psychology, Amherst College, USA
5. Institute of Child Development, University of Minnesota, USA
6. Department of Psychology, George Mason University, USA

PAPER

Abstract

Why are some young children consistently willing to believe what they are told even when it conflicts with first-hand experience? In this study, we investigated the possibility that this deference reflects an inability to inhibit a prepotent response. Over the course of several trials, 2.5- to 3.5-year-olds (N = 58) heard an adult contradict their report of a simple event they had both witnessed, and children were asked to resolve this discrepancy. Those who repeatedly deferred to the adult’s misleading testimony had more difficulty on an inhibitory control task involving spatial conflict than those who responded more skeptically. These results suggest that responding skeptically to testimony that conflicts with first-hand experience may be challenging for some young children because it requires inhibiting a normally appropriate bias to believe testimony.

Research highlights

• One of first studies to investigate the role of inhibitory control in how children respond to testimony.
• Young children who consistently favored testimony over first-hand experience when the two were in conflict showed lower inhibitory control abilities than those who favored experience over testimony.
• One reason some children may find it difficult to respond skeptically is because it involves inhibiting the normally appropriate expectation that what people say is true.

Introduction

The amount of knowledge we acquire from personal observation represents a relatively small fraction of what we actually know. For example, I know that humans have kidneys, even though I have never seen a human kidney. I know that dinosaurs roamed the earth 65 million years ago, even though I was not there at the time. How do we learn about things that lie outside of our own experience? The answer, of course, is that we can learn from other people – from what they do and say. For example, you may not know why the sky is blue. However, if I tell you that it is because molecules in the air scatter blue light more than red (and if you have no reason to doubt my credibility), then you too now know why the sky is blue. In short, we are not bound by our own experience; provided we are receptive to it, we can take advantage of other people’s knowledge, experience, and expertise (e.g. Csibra & Gergeley, 2006; Harris, 2012; Tomasello, 1999).

Being receptive to verbal information that other people provide (‘testimony’) makes it possible to learn all kinds of useful (and useless) things – from language to the fact that Thomas Jefferson wrote the Declaration of
Independence to the name of Kim Kardashian’s baby (“North West’). However, it also comes with a potential risk: Although the vast majority of testimony is likely to be true (e.g. Dennett, 1981; Grice, 1975), for a variety of reasons – including error, ignorance, and deception – people sometimes say things that are false. As a result, a learner cannot simply believe everything she or he is told.

One situation that could lead one to reject a piece of testimony is if that testimony conflicts with a belief formed on the basis of first-hand experience (e.g. Coady, 1992; Jaccard, 1981). For example, if someone claims that potatoes grow on trees, but you have seen that they grow underground, you would likely dismiss the testimony rather than doubting what you had seen. Knowledge obtained through first-hand experience is not infallible, but it certainly feels more salient, immediate, and reliable than information provided by someone else (e.g. Gelman, 2009; Locke, 1689/1964). Interestingly, however, some young children trust what they are told even when it conflicts with something they have just seen (Jaswal, 2010).

For example, in a study by Jaswal (2010, Study 3), 2.5-year-olds watched a goldfish cracker travel down a clear tube and land in one of three opaque cups. When they were invited to indicate where the cracker landed, they were usually correct. However, when an adult then asserted that it was in a different (wrong) cup, about half of the sample consistently responded in a deferential manner: On three or more of six trials, they searched the incorrect location mentioned by the adult even though they had earlier correctly identified the location where the cracker had fallen. (A number of additional conditions showed that children were not searching the named cup simply out of a desire to comply or please the adult, or because they had been asked the same question twice.) Given the young age of the participants, deference on one or two trials might not be surprising, as it could reflect inattention, curiosity about whether something could be found where the adult claimed, or confusion about the task. But consistent deference is surprising because there was a penalty each time a child ignored what s/he had just seen and responded deferentially – the loss of a tasty goldfish cracker. Why were some children consistently deferential while others were not?

One candidate for a characteristic on which the two groups may differ is inhibitory control. Perhaps the children who were consistently deferential had lower levels of inhibitory control compared to those who were more skeptical. This hypothesis follows from an influential proposal in the adult literature, which suggests that adults are biased to believe what they are told. Drawing on Spinoza, Gilbert (1991) has argued that the act of comprehension entails an initial acceptance of that which is comprehended. For example, continuing with the potato example from above, if someone tells you that potatoes grow on trees, you will initially accept this proposition as true even though you have seen that they grow underground. You can subsequently ‘unaccept’ testimony, but on Gilbert’s account, this is an effortful process; it requires overriding the default bias to believe. In line with this proposal, Gilbert, Krull and Malone (1990) found that when adults were operating under cognitive load, they were more credulous toward new information than when they were not under cognitive load (but see Hasson, Simmons & Todorov, 2005).

If children share with adults a default bias to believe testimony, then one reason some children in Jaswal (2010) were consistently deferential when faced with a conflict between first-hand experience and testimony could be that they had difficulty inhibiting this bias (for an account of how this bias may develop, see Jaswal, 2013; Jaswal, Croft, Setia & Cole, 2010). In the study reported here, we examined this possibility by investigating whether children who consistently deferred to testimony that conflicted with first-hand experience would experience more difficulty on an inhibitory control task than children who were more skeptical.

There has been one previous investigation of a possible link between inhibitory control and children’s trust in testimony. In Heyman, Sritanyaratana and Vanderbilt (2013), 3- and 4-year-olds were introduced to a ‘mean and tricky’ Big Bad Wolf puppet who repeatedly offered misleading advice about where he had hidden a sticker: He claimed that it was in one of two locations when it was always in the opposite. Children also participated in two executive function tasks: day-night (Gerstadt, Hong & Diamond, 1994) and head-to-toes (Ponitz, McClelland, Jewkes, Connor, Farris & Morrison, 2008), both of which required them to respond in a manner opposite to a given cue (e.g. respond ‘night’ to a picture of the sun; touch toes when the experimenter indicates to touch head). Over the course of four studies, Heyman et al. generally failed to find a significant correlation between how often children rejected the Big Bad Wolf’s bad advice and how well they performed on either of the executive function tasks. They concluded that children’s ability to distrust an ‘overtly misleading informant’ is independent of executive function (e.g. 2013, p. 654, p. 663).

Heyman et al.’s (2013) findings contradict our hypothesis. However, there are important differences between the two studies. For example, our interest is in individual differences in how children resolve conflicts between first-hand experience and testimony rather than individual differences in children’s ability to do the opposite of what someone says. In addition, our study
(which was designed and completed prior to learning of Heyman et al.) differs in that we were interested in how children respond to misleading testimony from a putatively trustworthy source rather than an explicitly deceptive informant. We will elaborate on the role these differences might play on outcomes involving inhibitory control in the Discussion.

In the study reported here, children between 2.5 and 3.5 years of age participated in the misleading testimony paradigm described earlier (Jaswal, 2010), in which children are repeatedly provided testimony that conflicts with an event they have just seen. We chose this age range because, based on that earlier work, we could be certain that about half of the sample would be consistently deferential. Our measure of inhibitory control was a spatial conflict task that required children to respond to a target on the basis of its identity. On the critical trials, this necessitated inhibition of the prepotent tendency to respond on the basis of the target’s spatial location (Gerardi-Coulton, 2000; Rothbart, Ellis, Rueda & Posner, 2003).

We chose the spatial conflict task as our measure of inhibitory control because it was simple to administer, understandable, and motivating to 2.5- to 3.5-year-olds. It is less familiar than some preschool measures of inhibitory control, but has a well-documented link to the construct (Gerardi-Coulton, 2000; Rothbart et al., 2003). For example, Gerardi-Coulton found that toddlers’ performance on the spatial conflict task was positively related to parent report of effortful control on the Children’s Behavior Questionnaire (Rothbart, Ahadi, Hershey & Fisher, 2001), as well as to their performance on a lab-based inhibitory control battery measuring the ability to delay gratification for an M&M or Cheerio, whisper the names of known cartoon characters, and take turns building a block tower (Kochanska, Murray, Jacques, Koenig & Vandengeest, 1996).

As noted, our primary interest in the current work was whether the consistently deferential children would find the inhibitory control challenge posed by the spatial conflict task more difficult than those children who were more skeptical. However, we also included three additional tasks for exploratory purposes: (1) a theory of mind scale (Wellman & Liu, 2004), included to investigate the possibility that children who are better able to understand that, for example, beliefs can be mistaken would respond more skeptically to the misleading testimony than children who do not share that understanding; (2) a newly designed ‘emotion-menu’ task, intended to assess children’s ability to respond skeptically to implausible testimony about emotions; and (3) a measure of emotion understanding (e.g. Denham, 1986), which was intended to help to interpret results from the emotion-menu task.

Method

Participants

Fifty-eight 2.5- to 3.5-year-olds (M = 34.76 months; SD = 3.01, range = 30 to 41 months; 27 boys, 31 girls) participated at one of two laboratories (n = 7 at one lab, and n = 51 at the other). Three additional children participated, but were not included because they failed to complete the misleading testimony and/or spatial conflict task. Most participants were White and from middle-class backgrounds, and all were exposed primarily to English in the home. There were no significant effects associated with testing location.

Design

Children participated in five tasks: misleading testimony, theory of mind, spatial conflict, ‘emotion-menu’, and emotion understanding. The misleading testimony task was always presented first, and the spatial conflict task was always presented last; the other three tasks were presented in random order. The emotion menu task turned out to be too difficult for the young children to understand, and they performed poorly. Thus, neither it nor the emotion understanding task (which we expected to use to interpret results from the emotion menu task) will be described further.

Measures and procedure

Children participated individually in a single visit, lasting between 30 and 45 minutes; they were encouraged to take breaks when needed.

Misleading testimony task

The misleading testimony task employed an apparatus modified from one designed to study spatial reasoning (Hood, 1995). The apparatus, shown in Figure 1, consisted of a wooden frame, measuring 61 cm high, 55 cm long, and 12 cm wide. Three ‘chimneys’ (each measuring 6 cm in diameter) were mounted at equal intervals into the top brace of the apparatus. Three opaque cups (each measuring 5 cm in diameter and 9 cm in height) could be placed into the lower brace, 47 cm beneath the chimneys. Each cup had affixed to it color stickers of typical exemplars from one of three categories: dogs, birds, or bears. A Plexiglas screen (53 cm by 16 cm) was mounted in front of the three cups to prevent children from being able to remove them from the apparatus. Each chimney could be connected to a non-adjacent cup using a flexible, clear tube (6 cm in diameter and approximately
60 cm in length). For example, the right-most chimney could be connected to the left-most cup.

We used the same procedure as Jaswal (2010). Children sat at a small table across from the experimenter and next to an adult confederate. They were first familiarized with each piece of the apparatus and introduced to the general requirements of the task. The experimenter introduced the three opaque cups, one at a time in random order. Children and the confederate were encouraged to hold each cup and to touch the stickers. Next, the experimenter placed all three cups in a row in the center of the table and asked children to point to each one in turn. For example, she asked, ‘Can you point to the bird cup?’ Children were given corrective feedback if necessary, and the confederate was asked to identify the same cup (and always did so correctly). These warm-up trials continued until children successfully identified the dog cup, the bear cup, and the bird cup in succession. The three cups were then removed from the table.

In the next part of the warm-up, the experimenter familiarized children with the goldfish crackers and clear tubes. She offered children (and the confederate) a cracker to eat, and then dropped one or two through a horizontally oriented clear tube, commenting on how a fish dropped in one end would come out the other. She then offered children and the confederate the opportunity to drop a fish or two through as well.

Finally, the apparatus was introduced, with the experimenter pointing out the three chimneys in the top brace and placing the three cups in position in the bottom brace. She held a goldfish cracker directly over one of the cups, dropped it in, and asked children to identify (by saying or pointing to) the cup in which it was now hidden. Children were always correct. After they had responded, she asked the same question of the confederate, who also always named the correct cup. She then turned back to the children and asked them to again identify the cup where the cracker had fallen. (This back-and-forth interaction was intended to familiarize children with the sequence of events that would take place during test trials.) Children always named the same (correct) cup they had earlier indicated. This sequence of events occurred three times, with the experimenter dropping the cracker in each cup once. She then attached each of the three clear tubes to the apparatus linking each chimney to a non-adjacent cup below in the configuration shown in Figure 1, and the test trials began.

Each child participated in six test trials. On each trial, the experimenter dropped a cracker through a chimney, and children and the confederate watched as it traveled down the attached clear tube and landed in an opaque cup (at which point the cracker could no longer be seen). The experimenter invited children to indicate (by pointing or saying) where the fish had landed (the Plexiglas barrier prevented them from directly grabbing the cup). As will be described below, this ‘pre-testimony response’ was usually correct. The experimenter then turned to the confederate (seated next to the child) and asked her where the cracker landed. The confederate offered misleading testimony, claiming that the cracker had landed in the cup directly beneath where it had been dropped. Finally, the experimenter returned her attention to the child and, with a puzzled expression on her face, asked him or her to resolve the conflict by making a ‘post-testimony’ response: ‘Where should we look?’

The experimenter removed the cup indicated by the child’s post-testimony response, and turned it upside-down. If the child indicated the correct cup, the goldfish cracker would fall out of the inverted cup, the experimenter would say, ‘You found it!’ and offer it to him or her to eat. If the child’s post-testimony response was incorrect, nothing would fall out and the experimenter would say, ‘Not there! Where should we look next?’ In this case, the child was allowed to continue indicating additional cups until the cracker was located; once located, it was put aside and not given to the child to eat. (At the end of the session, the experimenter did offer several goldfish crackers to the children.) Across the six test trials, the experimenter dropped the goldfish crackers into the chimneys in the same fixed order: left, right, middle, left, right, middle.

The session was videotaped and coded off-line. Coders noted the pre- and post-testimony locations indicated by

---

**Figure 1** Apparatus used in the misleading testimony task.
the child for each of the six trials. There were three locations to which children could make their pre- and post-testimony responses: the ‘correct’ location, corresponding to the cup where the ball actually landed; the ‘gravity’ location, corresponding to the cup directly beneath the chimney where the ball had been dropped (and the one to which the confederate’s misleading testimony referred); and the ‘other’ location, corresponding to the third cup on the apparatus. Each session was coded by two coders, who agreed on all trials.

Spatial conflict task

The spatial conflict task was modified from Gerardi-Caulton (2000) and Rothbart et al. (2003). Children sat at a table, next to the experimenter and approximately 15 inches in front of a 22- or 25.5-inch HP Touchsmart touchscreen monitor (monitors at the two testing locations differed in size).

Training began with a black screen showing a blinking red circle in the center, and the white outlines of a house in the lower left and lower right corners. Inside the left house was an image of Elmo, and inside the right house was an image of Grover. The experimenter named each character and explained that the left house belonged to Elmo (which is why his picture was inside it) and the right house belonged to Grover. She invited the children to touch each house.

Next, the experimenter explained that they were about to play a game, the goal of which was to help Sesame Street (or Muppet) characters find their way home. At this point, the blinking circle disappeared from the center of the screen and Elmo (or Grover) appeared in its place. The two houses (and their occupants) remained visible in the left and right corners of the screen. The experimenter pointed out Elmo in the center of the screen (the image was identical to the one at the bottom) and told children that touching his house would help him find his way home. When she touched Elmo’s house, the computer emitted a ‘ping’ sound, the Elmo in the center of the screen disappeared, and the Elmo inside the house bounced up and down for 2 seconds. The experimenter said, ‘Look! He’s happy because he found his house!’

The red blinking circle reappeared in the center of the screen, and the experimenter began the next training trial. As before, the circle in the center of the screen was replaced by a target image of Elmo or Grover, and the experimenter invited children to help him find his way home by touching the appropriate house. If children were correct, the experimenter responded enthusiastically, the computer emitted a ‘ping’ sound, the character in the center of the screen disappeared, and the character inside the touched house bounced up and down for 2 seconds. If children were incorrect, the experimenter reminded children of their error and of the correct response, the computer emitted a buzz, and the screen went dark for 2 seconds. After these first two training trials, children were encouraged to respond as quickly as they could (while maintaining accuracy) because the characters wanted to get home as fast as they could.1

There were eight training trials, four each with Elmo and Grover as the target. The target was randomly selected on each training trial without replacement, and the experimenter provided verbal feedback on each one. The left–right positions of Elmo and Grover at the bottom of the screen remained fixed throughout the training trials.

There were 32 test trials, which differed from the training trials in that the target appeared above a particular house rather than in the center of the screen. On 16 ‘compatible trials’, the target appeared above the target character’s house, and on 16 ‘incompatible trials’ it appeared above the opposite house. Children received the same computerized feedback as during training, with a ping and a bouncing of the target character in its house occurring for a correct response, and a buzz and screen darkening for an incorrect response. No verbal feedback was provided during test trials, but the experimenter did occasionally encourage children to continue.

The 32 test trials were divided into four blocks of eight trials each. Each block involved a different pair of characters: Bert/Ernie, Miss Piggy/Kermit, Big Bird/Cookie Monster, and Oscar/Zoe. Each character within a pair served as the target on four of the eight trials in that block; two of the trials involving each character as target were compatible trials, and two were incompatible. As a result of this counterbalancing, within each block, the correct response was to touch the left house on two compatible and two incompatible trials, and to touch the right house on two compatible and two incompatible trials. Which character in a pair served as a target on a given trial and whether that trial was compatible or incompatible was randomly determined (without replacement), with the constraint that no more than two compatible or incompatible trials could occur in a row in a given block. Within each block, the left–right position of the two response options remained fixed (i.e. Bert appeared in the left house and Ernie in the right house for all eight trials of the Bert/Ernie block).

---

1 We encouraged participants to place their hands on hand shapes affixed to the table between trials, as we had hoped to collect reaction time data and wanted their starting location to be the same on each trial. Unfortunately, many children did not do so before the program advanced to the next trial. Reaction time data were therefore not reliable and will not be presented here.
Data from the spatial conflict task were automatically collected as the task was being run, and consisted of the number of correct compatible and incompatible trials.

Theory of mind tasks

Children were presented with a modified version of Wellman and Liu’s (2004) theory of mind scale. Given the age of our youngest participants (2.5 years), we started with the task that Wellman and Liu’s Guttman scaling indicated was easiest (diverse desires) and proceeded through the more difficult tasks (diverse beliefs, knowledge access, false belief, hidden emotion) until the child failed a task. A child’s theory of mind score was the highest task number on which she or he succeeded (possible range = 0–5).

The theory of mind tasks were coded online during administration, and then double-coded from video by a separate coder. Four participants were mistakenly given a task one beyond where they had failed (they failed that one too). We used the lower score representing where testing should have stopped.

Results

Preliminary analyses failed to reveal any effects of sex or age in months, and so in the following analyses, we collapsed across these two variables.

Misleading testimony task

Figures 2a and 2b show how often children selected each of the three possible cups for their pre- and post-testimony responses during the misleading testimony task. As Figure 2a shows, children were clearly able to follow the goldfish cracker as it traveled down a clear tube and landed in one of the opaque cups: Their pre-testimony responses were correct, on average, on 5.41 ($SD = .96$) out of 6 trials. As Figure 2b shows, their post-testimony responses were correct significantly less often – on average, on 3.03 ($SD = 2.44$) trials out of 6, $t(57) = 7.85$, $p < .001$, $d = 1.03$. The reduction in selections of the correct cup in post-testimony responses was accounted for almost entirely by an increase in selections of the cup mentioned by the confederate (the gravity cup): The average number of gravity cup selections increased from 0.48 ($SD = .92$) of pre-testimony responses to 2.78 ($SD = 2.41$) of post-testimony responses, $t(57) = 7.80$, $p < .001$, $d = 1.02$.

The shading within each bar of Figure 2b shows the average number of correct, gravity, and ‘other’ pre-testimony responses that preceded a given post-testimony response. For example, of the 2.78 gravity post-testimony responses, 2.29 were preceded by a correct pre-testimony response, 0.45 were preceded by a gravity pre-testimony response, and 0.03 were preceded by an ‘other’ pre-testimony response. Clearly, as in Jaswal (2010), the confederate’s misleading testimony influenced the pattern of at least some children’s post-testimony responses.

Figure 2  (a) Average number of pre-testimony selections of the Correct (C), Gravity (G), and Other (O) cups in the misleading testimony task. (b) Average number of post-testimony selections of each cup. Shading within each function shows the pre-post conditional data (e.g. how many post-testimony gravity responses were preceded by pre-testimony correct or gravity or other responses). Error bars show SEM.

© 2014 John Wiley & Sons Ltd
One might be concerned that when children made different pre- and post-testimony responses, they did so for pragmatic reasons rather than because they were weighting the confederate’s testimony more heavily than their first-hand experience. That is, perhaps some children interpreted being asked the same question a second time as a signal that their pre-testimony response was incorrect (e.g. Siegal, Waters & Dinwiddy, 1988). In fact, previous work using this paradigm has shown that this pragmatic explanation does not apply in the current case. In Jaswal (2010, Study 4), children were asked to make a pre-testimony response, heard the confederate provide misleading testimony (as in the current study) or neutral testimony, and then were asked to make a post-testimony response. Results showed that children often made different pre- and post-testimony responses on trials involving misleading testimony (as in the current study), but never did so on trials involving neutral testimony. In other words, simply being asked the same question twice did not lead children to change their answers. Instead, the confederate’s misleading testimony was responsible.

Perhaps children were initially willing to believe the confederate, but after being repeatedly misled (and deprived of tasty goldfish crackers), they became more skeptical. In fact, this was not the case: Children selected the gravity cup on 1.47 (SD = 1.30) of the first three post-testimony trials, and 1.29 (SD = 1.24) of the final three post-testimony trials, t(57) = 1.49, p = .14. In short, as a group, children did not become more skeptical over the course of the six trials of the misleading testimony task.

Our measure of deference was based on how often children indicated the correct cup on their pre-testimony response and then switched to the (incorrect) cup mentioned by the confederate on their post-testimony response.2 Given the three possible response locations, the likelihood of this sequence of pre- and post-testimony responses occurring by chance was .11 (.33 * .33), or 11%. This pattern occurred, on average, on 2.29 (SD = 2.26) of the six trials (about 40% of the time), which is more often than would be expected by chance, t(57) = 5.49, p < .001, d = .72. As in Jaswal’s (2010) study, however, and as shown in Figure 3, there was considerable variability in how often individual children deferred, ranging from never (0) to always (6).

Figure 3  Distribution of deferential trials on the misleading testimony task. Deferential responses involved a correct pre-testimony response and a gravity post-testimony response.

Given a .11 probability of making a correct pre-testimony response followed by an incorrect post-testimony gravity response, a child who responded in this sequence on three or more of the six trials did so more often than expected by chance (binomial, p < .05). As Figure 3 shows, 25 (14 girls, 11 boys) deferred on three or more of the trials, and 33 children (17 girls, 16 boys) deferred on fewer than three trials. These will constitute, respectively, our ‘consistently deferential’ and ‘not consistently deferential’ groups (‘deferential’ and ‘skeptical’ for short) in the analyses that follow. The average age of the deferential group was 33.96 months (SD = 3.08), and the average age of the skeptical group was 35.36 months (SD = 2.85); the two groups did not differ in age, t(56) = 1.80, p > .05. Our interest was in whether differences in inhibitory control and/or theory of mind might distinguish these groups.

Spatial conflict task

As Figure 4 shows, consistently deferential children were more accurate on the compatible than the incompatible trials, whereas the skeptical children performed about equally accurately on both. A 2 x 2 mixed analysis of variance on these data (skeptical/deferential x compatible/incompatible trials) yielded no effect of whether children were in the skeptical or deferential group, F < 1, and a significant effect of the compatibility of the trials, F(1, 56) = 84.99, p < .001, ηp² = .21. This main effect must be interpreted, however, in light of a significant interaction between group and compatibility, F(1, 56) = 5.25, p = .03, ηp² = .09.

Simple effects analyses showed that the skeptical and deferential groups did not differ on the number of compatible trials correct, F(1, 56) = 1.32, p = .26, nor on

2 We could have instead analyzed a measure of skepticism, i.e. how often children chose the correct cup on both responses. This measure of skepticism and the measure of deference we chose to report were nearly perfectly correlated, r(56) = −.91, p < .001. That is, given a correct pre-testimony response, if children’s post-testimony response was not to the gravity cup (deference), it was almost always to the correct cup (skepticism). Thus, analyses on skepticism would simply be the mirror-image of those reported here.
the number of incompatible trials correct, $F(1, 56) < 1$. The skeptical group was equally accurate on compatible and incompatible trials, $F(1, 56) = 1.39, p = .24$. But the consistently deferential group was significantly more accurate on the compatible than incompatible trials, $F(1, 56) = 16.52, p < .001$, $\eta_p^2 = .22$.

We next calculated an interference score for each child, representing how much more difficult s/he found the incompatible trials to be relative to the compatible ones, controlling for performance on the compatible trials. Following Gerardi-Caulton (2000), we took the difference between the number of correct compatible and incompatible trials for a given child and divided by the number of compatible trials correct for that child. Thus, a child who was correct on 16 compatible and 8 incompatible trials would receive an interference score of 0.5, as would a child who was correct on 12 compatible and 6 incompatible. Although their overall level of performance differed (24/32 correct vs. 18/32 correct), the ratio of correct compatible to incompatible trials was the same. Interference scores could range from −1.0 to +1.0, with positive scores representing more difficulty on incompatible than compatible trials; negative scores representing the reverse; and scores close to 0 indicating a child who found the two types of trials to be equally easy (or difficult).

One way to analyze these data is to investigate whether there was a relation between the number of trials on which children deferred and their interference score. Neither of these variables was normally distributed (number of deferential trials: $D(58) = .20, p < .001$; interference score: $D(58) = .12, p = .04$), and so we used a Spearman’s rank order correlation, which was not significant, $r_s(56) = .165, p = .217$. This suggests that the extent of children’s deference was not predicted by the size of their inhibitory control score. However, as noted earlier, our primary interest was in investigating differences between children who were consistently deferential (i.e. those who deferred three or more times) and those who were not.

The average interference score of consistently deferential children was 0.19 ($SD = .32$), which is significantly greater than 0, $t(24) = 3.06, p = .005, d = .61$. Consistent with the accuracy analyses described above, these children had more difficulty on the incompatible than compatible spatial conflict trials. The interference score of the more skeptical children was 0.02 ($SD = .33$), which is not different from 0, $t(32) < 1$. As with the accuracy analyses above, skeptical children did not find the incompatible trials to be any more difficult than the compatible ones. Finally, the interference score of deferential children was significantly higher than that of skeptical children (0.19 vs. 0.02), $t(56) = 2.06, p = .04, d = .55$. In other words, consistently deferential children found the incompatible trials more challenging relative to the compatible ones than did children who were not consistently deferential.\footnote{We also performed the same analysis comparing children who never deferred with those who had deferred once or more. This analysis did not show a significant difference between these two groups. However, as noted in the text, our interest was in comparing the performance of children who were consistently deferential (i.e. those who deferred 3–6 times, more often than expected by chance) with those who were not consistently deferential.}

This same pattern of results can be seen at the individual level. Table 1 shows the distribution of interference scores for the consistently deferential and more skeptical groups. Most children in both groups had scores between −0.1 and 0.5. However, 5 of the 25 (20%) children in the deferential group had interference scores above 0.5, meaning that they were correct on fewer than

<table>
<thead>
<tr>
<th>Interference score</th>
<th>Deferential</th>
<th>Skeptical</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; −.50</td>
<td>0</td>
<td>3 (9%)</td>
</tr>
<tr>
<td>−.50 to −.11</td>
<td>4 (16%)</td>
<td>4 (12%)</td>
</tr>
<tr>
<td>−.11 to .10</td>
<td>7 (28%)</td>
<td>11 (33%)</td>
</tr>
<tr>
<td>.11 to .50</td>
<td>9 (36%)</td>
<td>15 (45%)</td>
</tr>
<tr>
<td>&gt; .50</td>
<td>5 (20%)</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Skeptical children deferred to the adult confederate’s testimony on fewer than half of the misleading testimony trials; deferential children did so on more than half. Interference scores could range from −1.0 to 1.0. Scores above 0 indicate more difficulty on incompatible than compatible trials; scores less than 0 indicate the reverse; scores around 0 show incompatible and compatible trials to be equally difficult.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Average number of correct compatible and incompatible trials on the spatial conflict task as a function of classification on the misleading testimony task. Error bars show SEM.}
\end{figure}

Table 1 Distribution of interference scores as a function of classification on misleading testimony task

© 2014 John Wiley & Sons Ltd
half as many incompatible as compatible trials. In contrast, none of the 33 children in the skeptical group had interference scores above 0.5.

A chi-square analysis on the data in Table 1 revealed a difference in the distribution of children’s interference scores in the skeptical and deferential groups, $\chi^2(4, N = 58) = 9.47, p = .05$. One could argue that the data from the negative end of the scale are difficult to interpret: Why would some children respond more accurately to the incompatible than compatible trials? To ensure that this subset of children was not responsible for the significant difference in distribution between the two groups, we conducted a second chi-square, this time only including children whose interference scores were close to or above 0 (i.e. $\geq -10$). Results showed that the distributions of interference scores in the two groups still differed, $\chi^2(3, N = 47) = 6.94, p = .03$.

**Theory of mind scale**

The deferential and skeptical groups did not differ in their average theory of mind scores: 0.92 ($SD = 0.86$) and 1.24 ($SD = 0.97$), respectively, $t(56) = 1.32, p = .19$. Although in principle, the theory of mind score can range from 0 to 5, in our sample, the highest score was a 3. As Table 2 shows, 20 children failed to respond correctly even to the first task (diverse desires), thereby receiving a score of 0; 14 succeeded on diverse desires, but failed on diverse beliefs, thereby receiving a score of 1; 22 succeeded on both diverse desires and diverse beliefs, but failed on knowledge access, thereby receiving a score of 2; and two succeeded on the first three tasks, but failed on false belief and so received a score of 3. The distribution of theory of mind scores did not differ by group, $\chi^2(3, N = 58) = 2.58, p = .46$. As we will discuss below, the restricted range of theory of mind scores obtained here makes it difficult to interpret this null result.

<table>
<thead>
<tr>
<th>Theory of mind score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deferential</td>
<td>10 (40%)</td>
<td>7 (28%)</td>
<td>8 (32%)</td>
<td>0</td>
</tr>
<tr>
<td>Skeptical</td>
<td>10 (30%)</td>
<td>7 (21%)</td>
<td>14 (42%)</td>
<td>2 (6%)</td>
</tr>
</tbody>
</table>

Note: Skeptical children deferred to the adult confederate’s testimony on fewer than half of the misleading testimony trials; deferential children did so on more than half. Theory of mind scores: 0 = failure on diverse desires; 1 = success on diverse desires, failure on diverse beliefs; 2 = success on diverse desires, diverse beliefs, failure on knowledge access; 3 = success on diverse desires, diverse beliefs, knowledge access, failure on false belief.

**Discussion**

Children who were consistently deferential to misleading testimony had more trouble navigating the inhibitory control challenge posed by the spatial conflict task compared to children who responded more skeptically. Consistently deferential children were less accurate on the incompatible trials, which required that they inhibit a prepotent response, than compatible ones, which did not. Skeptical children performed equally well on both types of trials.

Our results are consistent with the Spinozan account of belief outlined by Gilbert (1991), in which not believing testimony requires inhibiting a default bias to believe. In our study, children who could better inhibit a prepotent response were more likely to favor what they had seen in the misleading testimony task over what they were told. This finding, although consistent with the Spinozan account, is somewhat counter-intuitive: Given the qualitative feelings of immediacy and salience that accompany first-hand experience (e.g. Gelman, 2009; Locke, 1689/1964), one might have expected that belief in perceptual evidence would be the bias and that belief in testimony that contradicted that evidence would require cognitive effort (i.e. higher levels of inhibitory control in the deferential children than the skeptical ones). In fact, we found the reverse.

Our results differ from those obtained by Heyman *et al.* (2013). In four studies, they generally failed to find a significant correlation between how often 3- and 4-year-olds did the opposite of what a deceptive informant advised and performance on a pair of executive function tasks. As noted in the Introduction, there are important differences between how the link between inhibitory control and testimony was investigated in the two studies, which may have contributed to the conflicting conclusions.

First, Heyman *et al.* (2013) were interested in children’s rejection of testimony from an explicitly deceptive speaker. A Big Bad Wolf puppet was deliberately selected as the informant in that work so as to have negative associations for children. In addition, he was explicitly advised and performance on a pair of executive function tasks. As noted in the Introduction, there are important differences between how the link between inhibitory control and testimony was investigated in the two studies, which may have contributed to the conflicting conclusions.

First, Heyman *et al.* (2013) were interested in children’s rejection of testimony from an explicitly deceptive speaker. A Big Bad Wolf puppet was deliberately selected as the informant in that work so as to have negative associations for children. In addition, he was explicitly advised and performance on a pair of executive function tasks. As noted in the Introduction, there are important differences between how the link between inhibitory control and testimony was investigated in the two studies, which may have contributed to the conflicting conclusions.

First, Heyman *et al.* (2013) were interested in children’s rejection of testimony from an explicitly deceptive speaker. A Big Bad Wolf puppet was deliberately selected as the informant in that work so as to have negative associations for children. In addition, he was explicitly advised and performance on a pair of executive function tasks. As noted in the Introduction, there are important differences between how the link between inhibitory control and testimony was investigated in the two studies, which may have contributed to the conflicting conclusions.
inhibitory control plays in treating that testimony in a non-veridical way.

Second, the Big Bad Wolf in Heyman et al.’s (2013) studies always claimed that a hidden sticker was in one of two locations when it was actually in the opposite. Thus, although his testimony was false, it was actually informative in that it indicated where not to look (Mascaro & Sperber, 2009). In contrast, the testimony in the current work was false and uninformative: The confederate named an incorrect location but two possible locations remained. To respond correctly, children had to dismiss her testimony in favor of a belief formed on the basis of the event they witnessed earlier. Perhaps inhibitory control (as measured in our spatial conflict task) is more closely related to favoring perception over testimony than to doing the opposite of what someone says. Our study was not designed to compare an explicitly deceptive speaker to a putatively trustworthy one, or to compare different kinds of misleading testimony, and so future work will be required to investigate these possibilities.

One concern that could be raised about our misleading testimony task is that the confederate asserted that the goldfish cracker had landed in the cup directly beneath where it had been dropped. This location is consistent with the ‘gravity bias’, the expectation that young children have that unsupported objects will fall straight down. Using an apparatus similar to the one here but with opaque (rather than clear) tubes, Hood (1995) found that children younger than about 3.5 years of age searched for a ball in the cup directly beneath where it was dropped rather than following the trajectory of the opaque tube to the correct cup (see also, e.g. Bascandziev & Harris, 2010; Joh, Jaswal & Keen, 2011). Perhaps the children in our study who deferred to the adult testimony did so only because it was consistent with a pre-existing bias. If the testimony had not been consistent with a pre-existing bias, perhaps children would have been less likely to be deferential.

A related concern could be that responding skeptically on the misleading testimony task and correctly on the incompatible trials of the spatial conflict task both required that children refrain from making a ‘straight-down’ response. On the misleading testimony task, they had to refrain from selecting the cup that was directly beneath where the cracker had been dropped, and on incompatible spatial conflict trials, they had to refrain from selecting the response option that was directly beneath the target. If we had used measures of inhibitory control and misleading testimony that did not share this feature in common, perhaps we would not have found a link between the two.

We believe that the issues raised by these concerns are unlikely to have played a major role in our findings. First, Jaswal (2010) conducted a study (Study 2A) using the same apparatus and procedure, but the confederate’s misleading testimony alternated between indicating the ‘other’ incorrect cup and the gravity cup. The level of deference was the same in both cases: Whether the deferential response was straight down or offset from where the object was dropped did not make a difference. Thus, the particular (incorrect) location into which the confederate asserts the object has fallen does not appear to influence children’s skepticism or deference.

Second, although the inhibitory control task we used was spatial in nature, it has been linked to other, non-spatial measures of inhibitory control, including parent report of effortful control, as well as performance on a battery of other lab-based tasks (Gerardi-Caulton, 2000; Rothbart et al., 2003). Thus, we do not believe that our finding that consistently deferential children tended to have lower levels of inhibitory control is unique to the particular tasks we used. Nevertheless, it will be important for future work in this area to replicate and extend our finding using different inhibitory control and misleading testimony tasks.

We have argued that children are biased to trust testimony, and that those who had difficulty inhibiting this bias were the ones who consistently favored testimony over first-hand experience. There are, of course, factors other than inhibitory control that may influence how children respond to testimony. For example, we included a version of Wellman and Liu’s (2004) theory of mind scale to investigate the possibility that children who understood that beliefs could be mistaken would be more skeptical than those who did not share that understanding. As it turned out, however, there was not enough range in theory of mind scores to detect a difference. Indeed, only two of the children progressed to the point where they were given the false belief task, and neither of them passed. Thus, although there was no difference between the skeptical and consistently deferential groups in this particular study, we would not suggest that theory of mind is unrelated to how children evaluate testimony. With an older group of children or a more sensitive measure, a relation might have been obtained (see, for example, DiYanni & Kelemen, 2008; Fusaro & Harris, 2008; Vanderbilt, Liu & Heyman, 2011; but see Pasquini, Corriveau, Koenig & Harris, 2007).

Other work has identified several factors that play a role in how children respond to testimony. For example, using the same misleading testimony task as in the current study, Jaswal (2010) showed that children’s confidence plays an important role. Children who were given relevant experience with the apparatus before the confederate began
providing misleading testimony became skeptical of that testimony (presumably because the experience boosted their confidence in how the apparatus worked); those who were given irrelevant experience did not (see also Ma & Ganea, 2010). Jaswal and Malone (2007) found that preschoolers who heard an adult hedge before saying something surprising (e.g., ‘I think that’s a dog’ when referring to a cat-like animal) were more skeptical of the adult’s claims than those who did not hear the hedge. Certainly, a good deal of recent research in the selective trust literature has demonstrated that children as young as 2 years of age are more likely to endorse information from an informant who has been correct in the past than one who has been incorrect (for a review, see Harris, 2012). The strength of the bias to believe testimony (and therefore how much effort is required to respond skeptically) will clearly vary depending on a number of endogenous and exogenous factors (see Mills, 2013).

We have shown that children who consistently favor misleading testimony over first-hand experience had more difficulty on an inhibitory control task than those who were more skeptical of the testimony. We suspect that as the literature on individual differences in children respond to testimony grows, it will be messy and how much effort is required to respond skeptically) will clearly vary depending on a number of endogenous and exogenous factors (see Mills, 2013).

Acknowledgements

This research was supported by a grant from the Jacobs Foundation to VKJ and KPE. Thanks to the children who participated; to Johanna Kreafle, Claire Galdun, Katie Nelson, and research assistants for assistance in data collection and coding; and to Nameera Akhtar, Rebecca Dore, Marissa Drell, and Ashley Landrum for helpful comments on previous drafts.

References


Received: 6 August 2013
Accepted: 13 February 2014