ELSEVIER

Cognition 89 (2003) 25–41

COGNITION

www.elsevier.com/locate/COGNIT

Limits on theory of mind use in adults

Boaz Keysar[a,*], Shuhong Lin[a], Dale J. Barr[b]

[a]*The University of Chicago, Chicago, IL, USA*
[b]*The University of California, Riverside, CA, USA*

Received 14 September 2002; accepted 28 February 2003

Abstract

By 6 years, children have a sophisticated adult-like theory of mind that enables them not only to understand the actions of social agents in terms of underlying mental states, but also to distinguish between their own mental states and those of others. Despite this, we argue that even adults do not reliably use this sophisticated ability for the very purpose for which it is designed, to interpret the actions of others. In Experiment 1, a person who played the role of "director" in a communication game instructed a participant to move certain objects around in a grid. Before receiving instructions, participants hid an object in a bag, such that they but not the director would know its identity. Occasionally, the descriptions that the director used to refer to a mutually-visible object more closely matched the identity of the object hidden in the bag. Although they clearly knew that the director did not know the identity of the hidden object, they often took it as the referent of the director's description, sometimes even attempting to comply with the instruction by actually moving the bag itself. In Experiment 2 this occurred even when the participants believed that the director had a false belief about the identity of the hidden object, i.e. that she thought that a different object was in the bag. These results show a stark dissociation between an ability to reflectively distinguish one's own beliefs from others', and the routine deployment of this ability in interpreting the actions of others. We propose that this dissociation indicates that important elements of the adult's theory of mind are not fully incorporated into the human comprehension system.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Theory of mind; Perspective taking; Egocentrism

* Corresponding author. Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA. Tel.: +1-773-702-5830; fax: +1-773-702-0886.
E-mail address: boaz@uchicago.edu (B. Keysar).

0022-2860/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0010-0277(03)00064-7

1. Introduction

We understand our social world in mentalistic terms. John raising his arm is different from the metal arm lifting at the entrance to a parking lot. John could have raised his arm to catch attention, to catch a fly, to vote or to slap someone. We understand his action as goal driven, as motivated by a desire, a wish, an intent. We are able to conceive of such mental constructs and to distinguish them from physical constructs (e.g. Wellman, 1990). We are capable of telling apart our ambitions from our achievements, our desires from our reality. We are able to represent the world in pictures and in our minds and to conceive of these representations as separate from the world they represent. In this sense, we have a "theory" of mind (e.g. Olson, Astington, & Harris, 1988; Perner, 1988; Wellman, 1990).

This impressive human ability to hold a theory of mind has an important social function. Mental constructs provide explanations for our behavior and for the behavior of others. We use these constructs not only to understand actions but to predict the behavior of others. We can even use them to predict our own behavior as we "theorize" about what beliefs and desires we might have in future situations. Our theory of mind, then, allows us to navigate our personal and social world by explaining past behavior, and anticipating and predicting future actions (Moore & Frye, 1991).

Research on the child's theory of mind reveals a great deal about how young children develop the ability to reason about mental states. Presumably, young children have a rudimentary theory of mind which develops into an adult-like theory within a few years. Our focus is on the adult end of the continuum. We suggest that while adults have the ability to interpret social actions by means of a theory of mind, they do not exhibit the full-fledged theory of mind that is ascribed to them. Specifically, we argue that a major element of the theory of mind is not reliably applied by adults: adults' ability to represent others' beliefs is not reliably used to interpret others' behavior.

Our claim relies on the distinction between having a tool, and using the tool as part of one's routine operation. Here we focus on the ability to distinguish one's own beliefs from another's. By analogy, suppose you get an espresso machine as a gift. You already have a routine of making drip coffee, so if you replace the drip machine with the new machine you create a new routine. But you might decide to leave the new machine in the box and put it to use only when the need arises. When you need espresso, you take the machine out of the box, connect the parts and plug it in. Although it is true that children acquire this theory of mind machine by the age of 5 or 6 years at the latest, we argue that it is still "in the box" when they become adults. Though it could be used, it is not incorporated into the routine operation of the adult's system. Consequently, adults' use of crucial elements of theory of mind is not reliable.

1.1. The development of theory of mind in children: beliefs, desires and representations

Our claim is not about the use of mental constructs in general. It is reasonable to assume that a rudimentary theory of mind is indeed fully incorporated into our adult system. When the door to the supermarket opens as we step up to it, we see it as a mechanical action. When we raise our hand, the action is not perceived in mechanical terms. It is understood as an action in the service of an underlying mental motivation, a desire to stand out or an

intention to hurt. Children as young as 3 years old can appreciate the difference between mental and physical entities (Wellman & Estes, 1986). Young children realize that you can't touch a dream and that mental entities have different properties than physical entities (e.g. Estes, Wellman, & Woolley, 1989).

In fact, many basic elements of a theory of mind are present in very young children. Nine- to 12-month-old infants are already able to perceive an agent's action as goal oriented, perhaps even as intentional (e.g. Csibra, Gergely, Biro, Koos, & Brockbank, 1999; Gergely, Nadasdy, Csibra, & Biro, 1995; Woodward, 1998, 1999). Before the age of 2 years, children can have a meta-representation as is demonstrated in their ability to pretend play and to appreciate others' pretence (Leslie, 1994). Similarly, Bloom (2000) argues that children learn word meanings by figuring out the intentions behind speakers' referential expressions – again, demonstrating that the appreciation of intentionality appears very early in childhood.

Such a rudimentary theory of mind could be the basis for a more sophisticated understanding that goes beyond the appreciation of relations between beliefs and the reality they represent. Indeed, the hallmark of adult-like theory of mind is the ability to distinguish between one's own mental representations and those of others. It is the understanding that reality can be represented differently by different people. This is the crucial aspect of theory of mind that we focus on.

The literature on child development is divided on how early children develop the ability to treat representations separately from reality, and especially to distinguish between their own and others' beliefs. Some argue that this ability appears relatively early (e.g. Chandler, Fritz, & Hala, 1989; Fodor, 1992; Leslie, 2000) and others that it reliably appears only after the age of 4 (e.g. Gopnik & Wellman, 1992; Perner, 1991; Perner, Leekam & Wimmer, 1987; Wellman, 1990). But there is general consensus that by age 6 this ability is firmly in place. When young children start school their theory of mind is, for all intents and purposes, just like the adult's. Children show their sophisticated adult-like ability in two ways. They appreciate that others can be ignorant about something that they themselves know (Chandler & Greenspan, 1972; Marvin, Greenberg, & Mossler, 1976; Mossler, Marvin, & Greenberg, 1976) or that others can have false beliefs (e.g. Perner et al., 1987; Wimmer & Perner, 1983; Zaitchik, 1991).

The child's ability to appreciate false beliefs has been demonstrated in numerous studies. In their seminal paper, Wimmer and Perner (1983) presented young children with a situation that involved an agent with a false belief. For example, Maxi the puppet left chocolate in one cupboard and then left the room. In his absence, Maxi's mother moved it to another cupboard. So the child knew where the chocolate really was, but had reason to believe that Maxi's belief about its location was false. Three-year-old children tended to think that Maxi would later look for the chocolate where it really was, thus confusing Maxi's knowledge with their own. In contrast, 4- and 5-year-olds behaved more like grown adults. They thought Maxi would follow his own, false belief and look for the chocolate where he originally left it. They used Maxi's belief, not their own, to predict his action. A meta-analysis of many studies along these lines demonstrates that the developmental shift after age 3 holds across settings, countries, paradigms, and types of questions (Wellman, Cross, & Watson, 2001).

Highly related to the ability to appreciate others' false beliefs is the ability to appreciate

their ignorance or lack of belief. Indeed, several studies show that this ability develops at about the same time (e.g. Chandler & Greenspan, 1972; Marvin et al., 1976; Mossler et al., 1976) or somewhat earlier (e.g. O'Neill, 1996). For example, Mossler et al. showed children a video tape of a child requesting a cookie. Their own mother, who did not view the tape, then entered the room and the video was re-played but without the sound. The children knew that the child in the video was asking for a cookie, but would they realize that without the sound their mother cannot know that? Though 3-year-olds had difficulty, 4-year-olds were capable of attributing ignorance to their mothers. They were able to hold their privileged knowledge separate and to appreciate the fact that someone who does not have access to that information cannot use it. Given that such ability is firmly in place in pre-schoolers, our question is, what do adults do with this ability?

1.2. Adults' use of theory of mind: spontaneous vs. reflective

Our only concern is with the sophisticated element of theory of mind that enables adults to distinguish between their own beliefs and those of others, particularly in cases where others are ignorant or have a false belief. While we accept that children possess such an ability by age 6 at the latest, we argue that this ability is still "in the box" even for college students and beyond. Surprisingly, we find that adults fail to reliably deploy this ability precisely in the circumstances in which it would be most useful: when they interpret the actions of others. We argue that although adults can reflectively and deliberately use this sophisticated aspect of their theory of mind, this ability is not yet incorporated enough into the routine operation of the interpretation system to allow spontaneous, non-reflective use.

The evidence about the child's early ability to distinguish their beliefs from others' rests by and large on reflective tasks. Children are asked to consider where Maxi would look next, or they are asked to keep a secret from the experimenter and then asked if the experimenter knows the secret. Such questions tap a meta-cognitive ability, the child's ability to evaluate and reflect. To assess whether this ability is spontaneously used in understanding another's actions, we focus on adults' interpretation of another's linguistic behavior, a relatively less reflective domain.

Our current study builds upon a research paradigm reported in Keysar, Barr, Balin, and Brauner (2000), which addressed a different question pertaining to on-line linguistic processing. We therefore briefly describe the main finding in that study, and explain how our current studies are more directly relevant to the issue we are addressing here.

In the Keysar et al. study participants played a referential communication game with a "director". Several objects were put between the participant and the director in a free-standing grid, and the director gave instructions to move objects around in the grid. For example, the objects included a two-inch-high candle and a three-inch-high candle and the director said "Move the small candle to the right". While most objects were mutually visible, a few were occluded from the director and only visible to the participant. In this example, a one-inch-high candle was visible only to the participant. Despite the fact that the participants knew that the director was ignorant of the presence of the occluded small candle, adult participants often considered it as a referent. They sometimes reached for the one-inch candle, and they were delayed in identifying the intended object, the two-inch candle.

Keysar et al.'s study is directly relevant to an issue in the pragmatics of language use[1] but might be more limited in its application to the issue of theory of mind. The design is analogous to that of experiments that look at the child's ability to appreciate another person's ignorance. Just like in the experiments with children, adults in Keysar et al.'s experiment knew about the occluded candle and they knew that the director had no specific belief about the identity of occluded objects. Interestingly, adults did not use that knowledge to guide their interpretation of the director. But unlike the developmental studies that focus on conceptual perspective taking, Keysar et al.'s results could be explained as perceptually or visually induced. In general, perceptual perspective taking might not necessarily rely on the same mechanisms as conceptual perspective taking (e.g. Baron-Cohen, 1988; Langdon & Coltheart, 2001). Such a perceptually-based account reduces the relevance of those data for our current purposes.

A much stronger test of our hypothesis would use conceptual perspective taking. Consider the following situation illustrated in Fig. 1. An adult participant hides a roll of tape in an opaque paper bag. She now knows what is in the bag though she can no longer see it (Fig. 1A). She also knows that another participant, the director, does not know the identity of that object (Fig. 1B). The participant, then, is aware of the director's ignorance of the object that is in the bag. Given that the participant is an adult with a full-fledged theory of mind, she would have no difficulty distinguishing between her own knowledge and that of the director. For example, just like the 5-year-olds in the Marvin et al. (1976) experiment, she would easily provide a negative answer to the question "Does the director know what is in the bag?" But the question we raise is, would she spontaneously use this ability when she interprets the director's actions? Assume that the director tells her to "move the tape", referring to the cassette tape box which is mutually visible to both of them. Would she consider the hidden tape as the intended object, as Fig. 1C illustrates, despite her knowledge of the director's ignorance of the contents of the bag?

If theory of mind is fully incorporated into the operation of the adult system for interpreting others' actions, then adults' interpretations should be consistent with what they know the other knows and ignore anything that they know that the other doesn't know. Accordingly, if participants in our experiments are routinely guided by their theory of mind then they should not consider the hidden tape as the intended referent since they know that the director does not even know that it is there. If, however, the element of theory of mind that distinguishes between self and other's beliefs is not fully incorporated into the adult's system, then one would expect an incongruity between reflective measures and non-reflective measures. Thus, we predict that even though adults would have no problem assessing the director's ignorance, they would not reliably use this ability to arrive at the intention of the director. Therefore, we predict that the participant's private knowledge of the objects in the bag would be used when they search for referents for the director's instructions.

We report two experiments that test our hypothesis. We used a variety of measures in

[1] The standard pragmatic theory assumes that comprehension is restricted to mutual knowledge. One source of evidence for mutual knowledge is perceptual co-presence between speaker and addressee. The Keysar et al. (2000) experiment shows that perceptual co-presence does not guide comprehension because even perceptually-privileged objects were considered referents.

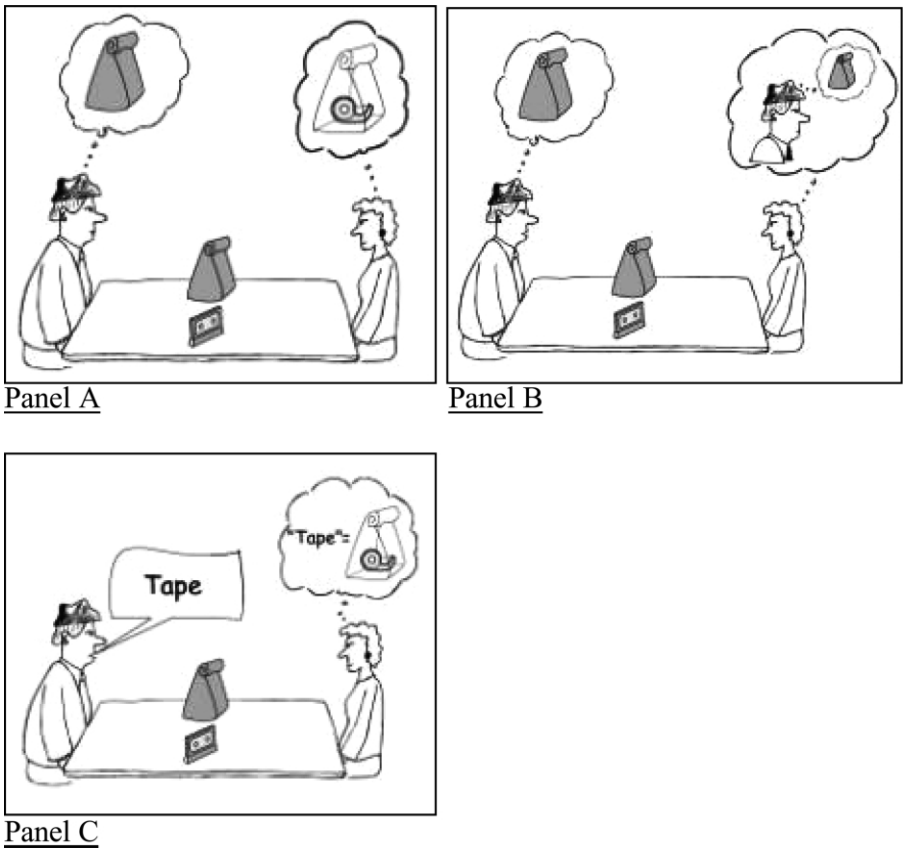

Fig. 1. *Who knows what*. A schematic representation of the participant's and the director's knowledge. While the participant knew what object was in the bag (A), she also knew that the director did not know (B). (C) shows that when the director says "the tape" the participant thinks he is talking about the hidden tape. Actual items included additional objects presented in a vertical array.

these experiments. The strongest, most important measure is behavioral: the tendency of participants to grab or try to move the bag. Whenever participants do that, they completely fail to use their knowledge of the director's beliefs about the contents of the bag. As a secondary measure we collected eye fixation data as an indication of which objects the participants consider as potential referents (Tanenhaus, Spivey- Knowlton, Eberhard, & Sedivy, 1995). Even when participants do not actually reach for the hidden object, their eye fixation pattern could reveal that they temporarily consider it as the target.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-eight native English speakers participated in the experiment. They were all

college students at the University of Chicago. None had a history of hearing or language disorders.

2.1.2. Materials and set-up

Participants sat at a table opposite a confederate director. On the table was a vertical array of 4 × 4 slots. Five of the slots were occluded from the director's perspective, and the remaining 11 were visible to both participant and director. One of these unoccluded slots included an object that was the target, such as a cassette tape box. Another object, such as a roll of tape, was hidden by the participant in a small brown paper bag and placed in an occluded slot. In addition to the intended object and the object in the bag, each array had several unrelated objects. For each grid, there was one "critical instruction" (e.g. "move the tape") in which the director gave an instruction to move a mutually-visible object that could also potentially refer to the object hidden in the bag.

The experiment had eight items, each with a different set of objects and critical instruction. Each item consisted of a series of instructions to move objects around and included one critical instruction. Each item also included one critical pair of objects, one of them the intended object and the other hidden in the bag, such as the cassette tape and the roll of tape.

To test our hypothesis we needed to detect any cases where participants failed to rely on their theory of mind. To do so we created situations in which participants would make a mistake if they did not consider the director's knowledge. So from the participant's perspective, without considering the director's ignorance, the hidden object was typically the best fit for the critical instructions. For example, the target for the instruction to move "the large measuring cup" was a medium size cup in the context of an even smaller measuring cup, both mutually visible to the director and the participant. Yet hidden in the bag was a larger cup. Therefore, when participants are not using their knowledge of the other's beliefs, they should consider the hidden object as the intended one. Note that while the hidden object was the best fit for the critical instructions, the intended object provided a perfectly good fit if one ignores the contents of the bag. We also considered and rejected the option of using an intended and a hidden object that were equally good referents because that would not have allowed us to test our hypothesis. For example, if the hidden measuring cup were the same size as the visible one, then the instructions "the large measuring cup" would have been ambiguous for participants who did not take the director's perspective into account. To resolve such ambiguity participants would have been forced to employ their theory of mind, thus obscuring the phenomenon we are attempting to uncover.

To collect baseline performance information for each item we added a condition in which the hidden object (e.g. roll of tape) was replaced with an object that did not fit the critical instruction (e.g. a battery). Thus, the "move the tape" instruction appeared after the participant had hidden in the bag either a roll of tape (experimental condition) or a battery (baseline condition). Each participant received half the items in the experimental condition and half in the baseline condition. Items and conditions were counterbalanced across participants. Order of presentation was random, with the provision that no more than two items in the same condition would appear consecutively.

A freestanding video camera recorded the scene on a HI8 recorder from behind and

above the participant, and the conversation was recorded on the same video tape. We also used an Applied Science Laboratories eye tracker to record the participant's eye movements. The participant wore a headband with a small camera lens, which filmed the left eye, and a magnetic head tracker provided information about head movement in space. Eye and head position information was integrated by computer to determine the gaze position. The participants were allowed to move their head freely so their natural interaction was not affected by the equipment.

A trained female confederate played the role of the director in order to ensure uniformity of critical instructions across conditions and participants. The confederate was well practiced in playing the role of a naive participant. To create a realistic situation, she indicated having some difficulty with the task, interjected her instructions with hesitations, and made occasional errors with non-critical objects. In addition, the director improvised most of the instructions, except that critical instructions for the target objects were scripted. Indeed, with the exception of one person, none of the participants later reported that they suspected during the experiment that the director was a confederate.

2.1.3. Procedure

The participant and the confederate arrived at the laboratory and the experimenter explained that they would be playing two different roles in a communication game. She then assigned roles, ostensibly randomly, and the participant received the role of the "addressee" while the confederate was assigned the director's role. At the beginning of each item, the director received a picture of the array of objects with arrows indicating where each object should be moved. The arrows were numbered to specify the order of object movement. The director then used this picture and instructed the participant to rearrange the objects accordingly.

The director's picture showed her perspective, meaning that only mutually-visible objects appeared on the picture, with the remainder of the slots clearly occluded. This was demonstrated to the participant, and the experimenter also pointed out that the objects in the occluded slots were not part of the game. In addition, before each item began, the experimenter put a large cardboard wall between the confederate and the participant as a visual barrier. Then she handed the participant an object and a brown paper bag and asked the participant to hide the object in the bag and place the bag in one of the occluded slots. The experimenter did not name the object but simply handed it to the participant and referred to it only as "this". After the participant had hidden the object in the bag and put the bag in the slot, the experimenter removed the barrier and the director started with the instructions.

The experiment began with a practice item to familiarize the participant with the task and to correct any misunderstanding. In order to make absolutely sure that the participant fully appreciated the director's difference in perspective, the participant and the director switched roles and the participant gave instructions for a second practice item. In this manner there would be no question that the participant understood the information provided in the picture of the array, appreciated that the director could not see hidden objects, and knew that the only objects relevant to the game were the mutually-visible ones. After the role reversal, the participant and the confederate resumed their original roles and the experimenter presented the first item. The experiment proceeded through all

eight items, with the director providing instructions and the participant moving the objects. Before each instruction the director said "ready?" at which point the participant looked at the center of the array and answered "ready". The participant was free to converse with the director, to ask questions and so on.

2.1.4. Design and predictions

Our design compares the experimental to the baseline condition within subjects, with each participant receiving four items in each condition. If participants' ability to distinguish between what they know and what others know guides their interpretation of the director's instructions, then they should consider only mutually-visible objects. Therefore, the experimental condition should not differ from baseline. In contrast, we argue that the ability to reason about the other's mental states is not fully incorporated into our system of spontaneously interpreting the actions of others. Therefore, we predict that participants would not reliably use this ability to understand the director's intentions. Instead, we predict that on occasion they would initially consider as intended referents objects that are known only to them; they might then recover and use their knowledge about the other's mind to correct their initial incorrect interpretation.

2.2. Results and discussion

Before reviewing the results, we wish to emphasize the simplicity of the participant's task. In each grid there was only one object that the director could not see and did not know about (with the exception of two items that had an additional, irrelevant object in an occluded slot). Moreover, the participant could not even see the object hidden in the bag. All that participants would need to do to successfully follow the director's instructions would be to ignore the existence of the one object that neither one of them could even see. Clearly, any consideration of the hidden object as a referent could only be due to a failure of conceptual perspective taking.

Our simplest behavioral measure provides the most stunning support for our hypothesis. Recall that each participant had four items in the experimental condition. In contrast to the baseline condition when participants never attempted to move the bag, in almost a third of the cases participants attempted to move the bag in the experimental condition (30%). The great majority of participants (71%) attempted to move the bag in at least one out of the four critical cases, and 46% attempted to move it for half or more of the items (see Table 1). They behaved as if they didn't know that the director was ignorant of the identity of the object in the bag. After they attempted to move the bag they either realized their mistake and self-corrected or the director corrected them.

Table 1
Percentage of participants who attempted to move the hidden object out of four possible items

	At least once	At least 50% of cases
Experimental	71	46
Baseline	0	0

Cases where participants grabbed the bag or reached for it are the clearest evidence that theory of mind is not reliably used for understanding the other. But participants could have considered the objects in the bag even when they did not reach for them. The eye fixation data provide a more sensitive, time-locked measure that can index comprehension in time. When people consider an object as a referent, their eyes quickly fixate on it (Tanenhaus et al., 1995). Given that participants tended to survey the array of objects during each trial, the baseline condition was useful in assessing the probability of looking at the bag when it did not contain a potential referent. For example, for the item with the hidden roll of tape we collected a baseline measure of eye fixations by hiding a battery in the bag instead of tape. Although participants in the baseline and experimental conditions saw the exact same set of objects at the moment of the critical instruction, we predicted that those who hid the tape would fixate on the bag more often and longer than those who hid a battery.

Participants' gaze position was represented by a cross-hair which was superimposed on the video tape. In addition, the spatial coordinates of the eye fixations over time were logged digitally. We counted a fixation on an object if the point of gaze remained in the object's slot for at least 100 ms consecutively. A coder who was blind to condition identified the end points of the critical instructions on the video tape, and a computer program used the digital information of eye fixation coordinates to determine the values of most of the fixation dependent measures.

We defined a temporal window of observation, starting from the instance at which the director began naming the object ("the tape") until the last fixation on the intended object before the participant reached for it. To determine whether participants considered objects hidden in the bag, we counted the number of times their eyes fixated on the bag throughout the observation window. Across all test items, the great majority of participants (92%) fixated on the bag at least once in the experimental condition. Overall, participants fixated on the occluded slot five times more often when it contained a roll of tape (experimental condition) than when it contained a battery (means = 1.38 and 0.24, respectively; $t(37) = 6.82$, $P < 0.0001$). The total amount of time spent fixating the bag was about six times longer for the experimental condition compared to baseline (means = 591 and 83 ms, respectively; $t(37) = 5.94$, $P < 0.0001$). So when participants knew that the bag had a roll of tape, and they heard instructions to move "the tape", they often considered the hidden tape even though they clearly knew that the director couldn't possibly know what was in the bag. However, their knowledge of the director's ignorance did help them to eventually identify the correct object.

Participants' knowledge of the contents of the bag interfered with their ability to identify the intended object in two ways. First, as we already demonstrated, it sometimes led them to attempt to move the bag. Secondly, it slowed down the identification of the intended object (see Fig. 2). We considered the first time the participant fixated on the intended object as an index of initially noticing the object. For 95% of the participants this initial noticing was delayed, with an average delay of 919 ms compared to baseline ($t(37) = 4.7$, $P < 0.0001$). The final fixation on the intended object right before reaching for it was considered the decision point. This decision was delayed for 82% of the participants, with a 2249 ms average delay compared to baseline ($t(37) = 7.9$, $P < 0.0001$). These two measures together define a decision window from first noticing the target to finally selecting it as the intended object. This decision window more than

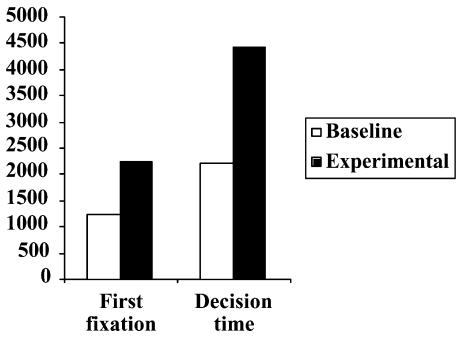

Fig. 2. *Delayed identification of the intended object*. Time in ms to notice the intended object ("First fixation") and to eventually decide that it is the intended object ("Decision time").

doubled from 985 ms in the baseline condition to 2315 ms in the experimental condition. The majority of participants (79%) showed such an enlarged decision window.

Clearly, our data show that participants often considered the object in the bag even though they knew that the director didn't share that information. Importantly, this pattern of data was typical for the great majority of participants and was not induced by outliers. Most participants attempted to move the bag at least once, all participants showed the predicted pattern on at least one eye fixation measure, and the great majority of participants (82%) showed the predicted pattern on at least three of the four fixation measures.

3. Experiment 2

The literature on the child's theory of mind considers two abilities in taking the other's perspective: appreciating another's ignorance, and appreciating another's false belief. The first experiment considers the case of ignorance because the participant believed that the director was ignorant about the contents of the bag. The second experiment generalizes the findings to the case of false beliefs and replicates the findings for the case of ignorance. The second experiment was similar to the first, except that in the false belief condition the experimenter misinformed the director about the contents of the bag in the presence of the participant. So in the tape example, when the participant hid the roll of tape, the experimenter showed the participant and the director a picture of a small leather toy ball which was supposedly in the bag. This created a false belief in the director and could have helped the participants to use their theory of mind for two reasons: (1) the false belief procedure highlights the difference between the two perspectives; and (2) reasoning about a concrete representation of an incorrect object might be easier than keeping track of the more abstract notion of lack of belief.

3.1. Method

Except for the false belief manipulation, the experiment was the same as the first experiment. Therefore, the method section will only highlight the differences.

3.1.1. Participants

Forty native English speakers from the University of Chicago participated in the experiment; 20 males and 20 females.

3.1.2. Design

Half the participants were in the ignorance condition and half were in the false belief condition. As in Experiment 1, half the items were in the experimental condition and the other half were in the baseline condition. The design was a mixed 2 (Hidden object: experimental vs. baseline) × 2 (Task: ignorance vs. false belief), with Hidden object as a within subject and Task as a between subject factor.

3.1.3. Materials and set-up

The materials were identical to Experiment 1, except for the pictures that were used to induce the impression of a false belief. The pictures were mounted on cards and showed small objects such as a small leather ball, a Japanese candy box and so on.

3.1.4. Procedure

The procedure replicated that of Experiment 1, except that in the false belief condition the experimenter showed to both the participant and the director a picture of the object that the participant supposedly hid in the bag. The picture was always different from the object the participant actually hid and could not have been a referent of the critical instructions. Participants were forewarned about the misinformation and cooperated in keeping it a secret.

3.2. Results and discussion

The data were coded and truncated exactly as in Experiment 1. In general, the results for the two tasks were identical. Participants attempted to move the bag reliably more often in the experimental condition than in baseline (24% vs. 0%) ($t(39) = 7.82$, $P < 0.001$). They did so to the same degree in the ignorance and the false belief conditions (26% and 22%, respectively) ($t(38) = 0.527$, $P = 0.6$). So even when participants believe that the director believes that the bag contains a small ball, they still attempt to move it when they hear "move the tape".

The eye movement data also presented a similar pattern for the two tasks. Overall participants fixated the bag much more often in the experimental than the baseline condition (means = 1.2 vs. 0.3) ($F(1, 38) = 53.44$, Mse = 0.299, $P < 0.001$). The effect held even for the false belief case (means = 1.1 vs. 0.3), and there was no interaction between Task and Hidden object ($F(1, 38) = 1.9$, Mse = 0.299, $P = 0.18$). Similarly, the time participants fixated on the bag was longer in the experimental than the baseline condition (means = 289 vs. 79 ms, respectively) ($F(1, 38) = 44.42$, Mse = 19799, $P < 0.001$), with no interaction between Task and Hidden object ($F < 1$). This shows that participants considered the hidden object both when they thought that the director was ignorant about it and when they thought that she had a false belief about it.

The hidden object also interfered with the identification of the intended object. Overall, noticing the intended object for the first time was delayed as the first fixation on it reveals

(means = 1969 vs. 1458 ms, in the experimental and baseline conditions) ($F(1, 38) = 8.46$, Mse = 616400, $P < 0.01$). Though participants were slower overall in the ignorance than in the false belief conditions, the difference between the experimental and baseline conditions was comparable (means = 2147 vs. 1699 ms, and 1790 vs. 1217 ms, respectively), yielding no interaction between Task and Hidden object ($F < 1$). Final fixation on the intended object was delayed by more than two seconds ($F(1, 38) = 26.83$, Mse = 3523127, $P < 0.001$), but similarly for the ignorance and false belief tasks (mean delay = 2244 and 2104 ms, respectively) ($F < 1$). As Fig. 3 illustrates, the decision window more than doubled in both the ignorance and the false belief conditions. The results of Experiment 2 clearly show that the effect from Experiment 1 generalizes from the case of ignorance to the case of false belief.

4. General discussion

Just like first graders, our adult participants had the ability to reflect upon the difference between what they and what the director knew. They were perfectly capable of saying that the director falsely believed that the object they themselves had just hidden in the bag was a small ball. Despite this reflective ability, they sometimes reached for the bag and were often delayed in identifying the intended object. When participants attempted to move the object that neither they themselves nor the director could see, they must have mentally selected the hidden object as the one intended by the director. This is particularly important because the ability to take the conceptual perspective of the other is an indispensable element in the fully-developed adult theory of mind. Our findings show that adults do not reliably consult this crucial knowledge about what others know when they interpret what others mean.

Some findings in the adult literature that show a tendency to impute one's own knowledge to others are similar to ours (e.g. Gilovich, Savitsky, & Medvec, 1998; Nickerson, 1999). But a closer look reveals that our findings are different. For example, the closest findings to our results might be Gilovich et al.'s discovery of the illusion among adults that their internal states are relatively transparent to observers. Participants in those

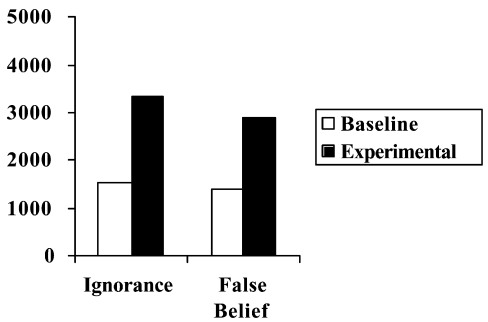

Fig. 3. *Delayed identification of the intended object as indexed by lengthening of the decision window*. Time in ms from noticing the target object to eventually deciding that it is the intended object as a function of condition for both ignorance and false belief tasks.

experiments tended to think that their feelings of disgust were discernible by others. Yet as Nickerson (1999) points out, the tendency to impute one's knowledge to others is "what one does in the absence of knowledge, or of a basis for inferring, that the other's knowledge is different from one's own" (p. 745). In contrast to such situations, our participants had firm knowledge of either the ignorance of the director or her concrete false belief. They knew exactly what she knew and didn't know and had no difficulty keeping the two apart. So despite the fact that our participants did not reflectively impute their own knowledge onto the director, they behaved as if the director was referring to an object of whose existence she was unaware. In short, they did not reliably deploy their sophisticated theory of mind to interpret the intention of the director.

Another related study might seem to contradict our claim, but we suggest that it is actually consistent with it. Nadig and Sedivy (2002) used a similar paradigm with 5- and 6-year-olds. They presented them with two glasses, one visible to both the speaker and the child, and one visible only to the child. When the speaker asked the children to move "the glass" they showed an early preference for the glass that the speaker could see. This suggests that considering the mind of the other could play a role from earlier stages of comprehension. This result is consistent with our claim for two reasons. While it is true that consideration of the other mind could come into play early, it does not happen reliably. Adults and children consistently show an egocentric component to comprehension (e.g. Keysar, 1994; Keysar & Barr, 2002; Keysar, Barr, Balin, & Paek, 1998; Keysar, Barr, & Horton, 1998). But more importantly, in the Nadig and Sedivy study the instruction "the glass" is ambiguous from the perspective of the child because either glass could be the intended one. Therefore, the child is forced to use theory of mind to identify a unique referent. We don't know if the children would have spontaneously used their theory of mind without such an ambiguity trigger. So while we show that people don't spontaneously use their theory of mind, Nadig and Sedivy show that children can use it quickly when they must.

Our argument should not be taken as a sweeping claim about theory of mind in general. Theory of mind has many elements, some of which are already in place at 2 years of age or even in infancy. For example, a rudimentary notion of intentional action is already clear with infants (Csibra et al., 1999; Gergely et al., 1995; Woodward, 1998, 1999), and children around the age of 2 are already able to talk about actions in terms of beliefs and desires. Our claim is specific to the ability to represent beliefs as separate from corresponding reality, thus allowing children to appreciate that different people can have mutually exclusive beliefs about the same reality, that some might be wrong and some ignorant about that reality. This is a sophisticated element of the theory of mind which develops relatively late (e.g. Wellman, 1990; Wimmer & Perner, 1983). It might very well be the case that early developing elements of the theory of mind spontaneously guide the interpretation of action; our results show that the later developing elements do not.

Some researchers argue that the ability to appreciate other's beliefs emerges before age 3, and that these younger children's ability is underestimated due to task demands (e.g. Bloom, 2000; Fodor, 1992; Leslie, 1994). Our findings are consistent with the idea that using this sophisticated ability is relatively difficult, but they add an important point. There is general agreement that by age 5 or 6 at the latest children are able to distinguish their

beliefs from others'. What we show is that the use of this ability is relatively unreliable even with adults.

Why might adults sometimes fail to deploy their fully-developed theory of mind? Perhaps in the "real world" perspectives tend to coincide such that what is present and salient to one person will tend to be salient to another. Under these circumstances, directly computing what another person knows or does not know at a given moment might be more trouble than it is worth. Furthermore, even when perspectives do not coincide, feedback from one's partner can obviate the need to compute that person's perspective for successful coordination. For example, although participants in our experiments often moved hidden objects, the director quickly corrected them and they eventually moved the correct object. Although participants could have pre-computed the director's perspective, they got away with being egocentric because they could count on the director's feedback. In short, the dynamic nature of face-to-face interaction gives people latitude to be egocentric by effectively distributing the burden of perspective taking across interlocutors (Barr & Keysar, in press).

Our findings might also shed light on two related issues: whether or not theory of mind is a specifically human ability, as well as the phylogeny of that ability in humans. It has been debated whether theory of mind is specific to humans (Povinelli, Bering, & Giambrone, 2000; Povinelli & Giambrone, 2001) or whether chimpanzees also show such ability (Premack & Woodruff, 1978). If indeed the sophisticated element of theory of mind is not fully incorporated into the cognitive system, it is reasonable to assume that it represents a relatively late addition to the human cognitive repertoire. The later such ability appears in humans, the more likely it is to be human specific. More particularly, it is possible that the ontogenously late-appearing ability to distinguish the mind of self and other is specific to humans, but is yet to be fully incorporated into our system.

Recall that a central function of a theory of mind is to predict and interpret the actions of self and others. In light of this function, it is curious that adults do not reliably use what they know other people believe in order to understand what they mean. Our findings, then, show a clear dissociation between an ability that is firmly in place by adulthood, and the reliable use of this ability for the very purpose for which it is designed.

Acknowledgements

We thank Jessica Jalbrzikowski, Shaun Kenny, Elizabeth Medvedovsky, Colleen Smith and Ruth Zaijcek for technical help and Nick Epley and Linda Ginzel for comments on an earlier version. This work was supported by PHS grant R01 MH49685-06A1 to Boaz Keysar.

References

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: cognitive or affective? *Journal of Autism and Developmental Disorders*, *18*, 379–402.

Barr, D. J., & Keysar, B. (in press). Making sense of how we make sense: The paradox of egocentrism in language

use. In H.L. Colston & A.N. Katz (Eds.), Figurative language comprehension: Social and cultural influences. Mahwaw, NJ: Erlbaum.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: Cambridge University Press.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind. *Child Development*, *60*, 1263–1277.

Chandler, M. J., & Greenspan, S. (1972). Ersatz egocentrism: a reply to H. Borke. *Developmental Psychology*, *7*, 104–106.

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*, 237–267.

Estes, D., Wellman, H. M., & Woolley, J. D. (1989). Children's understanding of mental phenomena. In H. Reese (Ed.), *Advances in child development and behavior*. New York: Academic Press.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*, *44*, 283–296.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, *75*, 332–346.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind is really a theory. *Mind and Language*, *7*, 145–171.

Keysar, B. (1994). The illusory transparency of intention: linguistic perspective taking in text. *Cognitive Psychology*, *26*, 165–208.

Keysar, B., & Barr, D. J. (2002). Self anchoring in conversation: why language users do not do what they "should". In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment* (pp. 150–166). New York: Cambridge University Press.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Sciences*, *11*, 32–38.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1–20.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: insights from a processing approach. *Current Directions in Psychological Sciences*, *7*, 46–50.

Langdon, R., & Coltheart, M. (2001). Visual perspective-taking and schizotypy: evidence for a simulation-based account of mentalizing in normal adults. *Cognition*, *82*, 1–26.

Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, *50*, 211–238.

Leslie, A. (2000). How to acquire a 'representational theory of mind'. In D. Sperber & S. Davies (Eds.), Metarepresentation, Oxford: Oxford University Press.

Marvin, R. S., Greenberg, M. T., & Mossler, D. G. (1976). The early development of conceptual perspective taking: distinguishing among multiple perspectives. *Child Development*, *47*, 511–514.

Moore, C., & Frye, D. (1991). The acquisition and utility of theories of mind. In D. Frye, & C. Moore (Eds.), *Children's theories of mind* (pp. 1–14). Hillsdale, NJ: Erlbaum.

Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2- to 6-year-old children. *Developmental Psychology*, *12*, 85–86.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence for perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329–336.

Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737–760.

Olson, D. R., Astington, J. W., & Harris, P. L. (1988). Introduction. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 1–15). New York: Cambridge University Press.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–677.

Perner, J. (1988). Developing semantics for theories of mind: from propositional attitudes to mental representations. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind*. New York: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., & Wimmer, H. (1987). Young children's understanding of belief and communicative intention. *Pakistan Journal of Psychological Research*, *2*, 12–40.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-old's difficulty with false belief: The case for a conceptual deficit. *British Journal of Development Psychology, 5*, 125–137.

Povinelli, D. J., Bering, J. M., & Giambrone, S. (2000). Toward a science of other minds: escaping the argument by analogy. *Cognitive Science*, *24*, 509–541.

Povinelli, D. J., & Giambrone, S. (2001). Reasoning about beliefs: a human specialization? *Child Development*, *72*, 691–695.

Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*, 655–684.

Wellman, H. M., & Estes, D. (1986). Early understanding of mental entities: a reexamination of childhood realism. *Child Development*, *57*, 910–923.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non purposeful behaviors. *Infant Behavior & Development*, *22*, 145–160.

Zaitchik, D. (1991). Is only seeing really believing? Sources of true belief in the true belief task. *Cognitive Development*, *9*, 91–103.