RESEARCH REPORT

Reasoning About Other People’s Beliefs: Bilinguals Have an Advantage

Paula Rubio-Fernández and Sam Glucksberg
Princeton University

Bilingualism can have widespread cognitive effects. In this article we investigate whether bilingualism might have an effect on adults’ abilities to reason about other people’s beliefs. In particular, we tested whether bilingual adults might have an advantage over monolingual adults in false-belief reasoning analogous to the advantage that has been observed with bilingual children. Using a traditional false-belief task coupled with an eye-tracking technique, we found that adults in general suffer interference from their own perspective when reasoning about other people’s beliefs. However, bilinguals are reliably less susceptible to this egocentric bias than are monolinguals. Moreover, performance on the false-belief task significantly correlated with performance on an executive control task. We argue that bilinguals’ early sociolinguistic sensitivity and enhanced executive control may account for their advantage in false-belief reasoning.

Keywords: bilingual cognition, theory of mind, false-belief reasoning, executive control, eye tracking

Experience can have powerful effects on cognitive abilities. For example, London taxi drivers have enlarged regions of the hippocampus involved in spatial navigation, and the extent of this enlargement correlates positively with the time spent taxi driving (Maguire et al., 2003). Action video-game playing enhances visual selective attention (Green & Bavelier, 2003). Canadian postal workers who regularly process letters and digits together in postal codes have better integration of letter and digit processing than do people without this experience (Polk & Farah, 1998). Experience with a variety of cognitive tasks across the life span increases cerebral reserve, reflecting the protective effect of intellectual exercise against cognitive decline with aging (Valenzuela & Sachdev, 2006).

Bilingualism provides experiences that can lead to various cognitive effects but with mixed outcomes (Bialystok, 2009; Bialystok & Craik, 2010). On the negative side, bilingual children have a smaller vocabulary in each language than do their monolingual peers (Oller & Eilers, 2002), and bilingual adults underperform in word retrieval tasks relative to monolinguals, partly because of the interference of their other language (Kaushanskaya & Marian, 2007). On the positive side, bilingual language production requires constant monitoring of the target language in order to minimize interference from the competing language. This requires the exercise of executive control, which in turn should strengthen a person’s executive control system (Bialystok, 2009; Bialystok & Craik, 2010; cf. Costa & Santesteban, 2004, for an alternative model of bilingual speech production without inhibitory control). Bilinguals’ advantage in executive control tasks has been found in children as well as in young and older adults (Bialystok, Craik, Klein, & Viswanathan, 2004; Carlson & Meltzoff, 2008; Costa, Hernández, & Sebastián-Gallés, 2008).

Bilingual children also perform better in false-belief tasks than do their monolingual peers (Goetz, 2003; Kovács, 2009). False-belief tasks have been a key test of social cognition development—as well as being used in the diagnosis of autism and related disorders—for more than 2 decades (Onishi & Baillargeon, 2005; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). In the classic Sally–Anne task (Baron-Cohen, Leslie, & Frith, 1985), children are presented with two puppets, Sally and Anne, who are playing with a toy. When they finish playing, they put the toy into a box and Anne leaves the scene. While she is away, Sally moves the toy to a different box. When Anne comes back, the child is asked where she will look for the toy. Monolingual children around the age of 4 years typically answer correctly that she will look for the toy in the original container. However, younger monolingual children as well as some autistic individuals tend to respond according to their own knowledge of the situation, thus failing to show an appreciation of Anne’s false belief about the location of the toy. In contrast, bilingual children as young as 3 years have shown a precocious success in false-belief tasks (Goetz, 2003; Kovács, 2009).
The extent to which young children may fail traditional false-belief tasks because they are not able to appreciate another person’s beliefs or because they are not able to inhibit their own knowledge of the situation has been a recurrent theme in the developmental literature (Perner & Lang, 1999; Wellman et al., 2001). Bilingual children’s advantage in traditional false-belief tasks might well be related to both factors. Bilingual children must develop an early sociolinguistic sensitivity to the language knowledge of their interlocutors because they must use their languages accordingly (Genesee, Boivin, & Nicoladis, 1996; Genesee, Nicoladis, & Paradis, 1995). Even though this fundamental aspect of bilingual children’s experience has not been investigated in the context of false-belief reasoning, their awareness that other people do not always speak the same languages as they do might be an early form of appreciating that other people might have a different perspective from their own. Moreover, this early form of perspective taking is combined with an early developed executive control system that is necessary to focus on the target language and avoid interference from the contextually inappropriate linguistic system (Kovács & Mehler, 2009). Their advanced executive control would help bilingual children inhibit their own perspective in false-belief tasks (Kovács, 2009). In general, language learning itself can promote the development of children’s theory of mind—their understanding of other people’s mental states and intentions (Pyers & Senghas, 2009).

Although it has been established that at around 4 years of age children are able to pass standard false-belief tasks (Wellman et al., 2001), more recent studies using eye-gaze measures to monitor infants’ expectations have suggested that even 15-month-old infants might be able to succeed in nonverbal versions of the false-belief task (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007). Nevertheless, although infants might be able to pass nonverbal versions of the false-belief task and older children perform optimally in standard false-belief tasks, adults’ performance on a variety of judgment tasks can still be affected by the curse of knowledge—a tendency to be biased by their own knowledge of the situation (Birch & Bloom, 2007; Camerer, Lowenstein, & Weber, 1998; Kelley & Jacoby, 1996; Keysar, 1994). It is therefore possible that, using sufficiently fine grained measures, adults might reveal an egocentric bias in their performance on a standard false-belief task.

We investigated the false-belief reasoning abilities of adult bilinguals relative to monolinguals using a standard false-belief task coupled with an eye-tracking technique. Monitoring participants’ eye movements during performance in the traditional Sally–Anne task should enable us to observe the effects of the curse of knowledge in adult false-belief reasoning. If adult participants first consider the container where they know the toy is before taking the perspective of the mistaken story character, then this would implicate an egocentric bias. Recent eye-tracking research has shown that adult participants do suffer from an egocentric bias when performing referential communication tasks (Epley, Morewedge, & Keysar, 2004; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). For example, in following the instruction “Move the small candle,” participants tend to initially consider the smallest of three candles on display, even though that particular candle is not visible to the speaker and therefore cannot be the intended target.

If eye movements reveal an egocentric bias in adult false-belief reasoning, do bilingual participants suffer less from the curse of knowledge than do monolinguals? Given their sociolinguistic sensitivity to the language background of their interlocutors as well as their enhanced executive control, bilingual adults may maintain an advantage in false-belief reasoning analogous to the advantage found in bilingual children (Goetz, 2003; Kovács, 2009).

**Method**

**Participants**

We tested 46 undergraduates at Princeton University: 23 bilinguals and 23 monolinguals. All participants were to some extent familiar with a second language. For the purpose of the study, we were interested in distinguishing those participants who had been regularly using two languages for most of their lives from those who had not. The two main criteria for classifying our participants as bilinguals or monolinguals were (a) how many years they had known a second language and (b) how regularly they had had to switch between their first and second languages in that time. To obtain this information, we used the Language History Questionnaire (Gullberg & Indefrey, 2003).

We established that the bilinguals in our sample should have learned a second language (L2) before age 9 and have been using it regularly for 10 years or more. The mean age of acquisition of the L2 was 3, and the mean number of years of use of the L2 was 16. The bilinguals in our sample had learned their L2 in (a) a bilingual household (12 participants), (b) a monolingual household using a language foreign to the country of residence (six participants), or (c) a bilingual school (five participants). The bilinguals in our group would have therefore had to switch languages daily or almost daily for a number of years. The L2s in our sample were Spanish (5), Chinese (8), Hindi (1), Korean (2), Russian (1), Hebrew (2), French (2), Farsi (1), and Japanese (1).1

Because all participants in the study were Princeton students, both bilingual and monolingual participants were recruited from a selective pool of high-performance students who had been accepted into university under rigorous academic criteria. Given our samples, no important differences in IQ or verbal ability would be expected. The mean age was 19.7 years for the bilinguals and 19.4

---

1 Having a collectivistic cultural background might give people an advantage in perspective taking (Wu & Keysar, 2007). However, this advantage seems to be observable only in individuals who were born and brought up in collectivistic countries. Wu and Keysar (2007) failed to find this effect in American Chinese students, whereas an advantage was found when comparing Chinese students who had arrived in the United States within the last year and American students (B. Keysar, personal communication, April 24, 2009). In our sample, 12 bilingual participants had an Asian collectivistic background, although five of those participants had been born in the United States. Three participants in our monolingual group also had an Asian collectivistic background. It was not originally our intention to investigate the effects of cultural background on false-belief reasoning. However, given the considerable number of Asian participants in our sample, we decided to further analyze our data comparing Asian collectivistic participants (n = 15) and non-Asian individualistic participants (n = 31) alongside the main comparisons between monolinguals and bilinguals.
years for the monolinguals. Of the 23 bilingual participants, 15 were female; of the 23 monolinguals, 12 were female. All participants had either been born in the United States (65% of bilinguals and 87% of monolinguals) or moved to the United States or to another English-speaking country by age 12. Crucially, at the time they took part in the study, English was the dominant language of all participants, whether bilingual or monolingual.

**Sally–Anne Task**

**Procedure.** The false-belief task was an extended computer version of the classic Sally–Anne task. Two kindergarten characters, Sally and Anne, interacted in an animated cartoon. Each child had a favorite toy and a container in which she kept it before going home every day. Participants were familiarized with the setting of the story in two warm-up trials before they were tested in the following two different conditions:

**False-belief (FB) condition:** Anne puts her doll in her basket and goes home. While she is away, Sally moves Anne’s doll from the basket to the box. When Anne comes back the next day, participants are asked “Where will she look for her doll?”

**True-belief (TB) baseline:** Sally puts her horse in her box and goes home. Anne was also going to put her doll in her basket, but because her basket is getting full, she decides to put it in the box. When Anne comes back the next day, participants are asked “Where will she remember to go and find her doll?”

The FB condition is the standard Sally–Anne scenario in which Anne erroneously believes that her doll is in the basket where she left it. The TB condition served as a baseline for FB because Anne—like the participants—is correct about the location of her doll. Because the TB baseline was always presented after the FB condition, we used a different comprehension question in the TB condition in order to avoid facilitation due to repetition of the question. Given the greater syntactic complexity of the TB question, however, longer response times (RTs) were expected in the baseline condition. Because predictability was a potential issue for finding evidence of an egocentric bias with a simple task, participants were tested in only one trial per condition, with the first experimental trial always corresponding with the FB condition. In order to separate the trials clearly, a short clip with an adaptation of a traditional children’s song was played after each scene.

Unlike the young children for whom the task was originally designed, adult participants were expected to respond to the FB question correctly. Any evidence of the curse of knowledge in this population should be observed only in their eye movements and RTs.

Participants were given standard instructions that described their role as a control group in an experiment aimed at children. Participants watched the Sally–Anne cartoon on a computer screen and listened to the accompanying story, which lasted approximately 4 min. At different points in the story they were asked a comprehension question, which they answered using one of two labeled keys on the computer keyboard. The position of the box and basket keys was symmetric to that of the containers on the screen. In order to avoid any possible bias related to a differential left- and right-hand dominance, participants were asked to press the response keys with their dominant hand.

In order to maximize the extent of eye movements, immediately before each comprehension question the containers disappeared from the scene while the narrative continued. Then Anne reappeared on the lower center of the screen to make sure that all participants were fixating on the same point when the two containers reappeared in the top corners of the screen (see Figure 1). The presentation of the display where the containers reappear on the screen was taken as the onset for eye fixations. The containers reappeared at the onset of the verb in the questions in order to obtain an accurate measure of the initial stages of perspective taking. A fixation was defined as an eye movement that remained on one of the target areas of interest for a minimum of 100 ms. The onset for RTs corresponded with the offset of the question.

**Results.**

**Gaze direction.** The first measure that we used was direction of first fixation in the FB condition. That is, on which target participants first fixated when the containers reappeared on the screen during the FB question. Overall, 19 participants (41.3%) looked first at the correct container where the child had left her toy, whereas 26 participants (56.5%) revealed an egocentric bias, looking first at the container where they knew the toy was before correcting this tendency and focusing their gaze on the correct container. One monolingual participant did not fixate on either container before responding to the question.

Considering bilinguals and monolinguals separately, 13 bilingual participants (56.5%) gazed directly at the correct container, whereas only six monolinguals (26.1%) did so, with the majority of monolingual participants revealing an egocentric bias (see Figure 2). This difference is reliable, \( \chi^2(1, N = 45) = 3.94, p < .048 \).

Whereas in the FB condition the wrong container represented the participants’ privileged knowledge of the location of the toy and the correct container represented the character’s perspective, in the TB baseline there was no competition between the two targets. It is interesting that in the TB condition, 21 bilingual participants (91.3%) and 18 monolingual participants (78.3%) first fixated on the correct target, which did not result in a reliable difference, \( \chi^2(1, N = 46) = 1.52, p = .218 \).

**Fixation latency.** The second eye-tracking measure was fixation latencies on the correct container in the FB condition. That is, how long it took participants to first fixate on the correct target when the containers reappeared on the screen at the end of the FB question. Data points that were above 2.5 standard deviations over the group’s mean fixation latency on the correct target were discarded as outliers. Outlying responses (two in the bilingual data set and two in the monolingual data set) as well as data points that did not reveal a fixation on the correct target (one in the bilingual data set and three in the monolingual data set) were replaced by the group’s mean fixation latency on the correct target without the outlying responses.

Given the larger number of first fixations on the wrong container among monolingual participants, bilingual participants were on average faster at first fixating on the correct target (mean 619 ms) than were the monolingual participants (mean 812 ms); see

---

2 Given the syntactic ambiguity of the TB question, the intended reading was stressed in the recording of the materials.
This difference was reliable, $t(44) = 2.07, p < .045$, thus confirming the better performance of the bilingual group.

Response time. The third measure was RTs for the FB and TB questions. Data points that were above 2.5 standard deviations above the group’s mean RTs for each condition were discarded as outliers. Outlying responses (one bilingual) as well as data points that revealed a second press on the response key (i.e., the participant had first pressed the response key before the offset of the question; one bilingual and one monolingual) were replaced by the group’s mean RT in that condition without the outlying responses.

Given that we used a different question for the FB condition and the TB baseline— with the second question eliciting much longer RTs than the first one because of its greater syntactic complexity (see Figure 4)— we computed the relative difference between participants’ RTs to the TB question and the FB question over the TB baseline. This relative measure of RT did not reveal a reliable difference between the two groups, $t(44) = 1.52, p = .135$, suggesting that RTs might be less sensitive than eye-movement measures.

Regarding baseline performance, RTs to the TB question did not differ reliably for bilinguals ($M = 1,692 \text{ ms}$) and monolinguals ($M = 1,563 \text{ ms}$), $t(44) = 1.38, p = .175$. We take the bilinguals’ comparable RTs in the TB baseline as evidence that this group was not overall superior to the monolingual group in the context of this experiment.$^3$

RTs to the FB question did not differ reliably for bilinguals ($M = 916 \text{ ms}$) and monolinguals ($M = 1,042 \text{ ms}$), $t(44) = 0.891, p = .378$. RTs therefore seem to be less sensitive than eye-tracking measures, which did reveal a reliable difference between bilinguals and monolinguals in the FB condition. When we collapsed the RTs for bilinguals and monolinguals and divided them between those who first fixated on the correct container ($M = 812 \text{ ms}$) and those who did not ($M = 1,097 \text{ ms}$), we also observed a reliable difference, $t(44) = 2.05, p < .048$. Given that this pattern of RTs was driven by accuracy of first fixation and significantly more bilinguals than monolinguals first fixated on the correct container, it is likely that with a larger sample of participants, a reliable difference in favor of the bilingual group might also be observed in the RT measure.$^4$

Simon Task

Procedure. After completing the Sally–Anne task, participants were given a version of the Simon task in order to assess their level of executive control (O’Leary & Barber, 1993). In this task, participants had to press a right-hand key when they saw the word RIGHT on the computer screen and a left-hand key when they saw the word LEFT. In the baseline condition, the location of the words on the computer screen matched the actual words

---

$^3$ It has been documented that English–Spanish bilinguals, for example, might show parsing preferences when processing certain syntactically ambiguous constructions because of interference from their other language (see Desmet & Duyck, 2007, for a review). This could suggest that at least some of the bilinguals in our sample might have had problems parsing the TB question—even though the intended reading was stressed in the recording. However, recent studies have shown that the erosion of parsing strategies in the first language occurs only after prolonged immersion in an L2 environment (Dussias, 2004; Dussias & Sagarr, 2007). Because all bilingual participants in our study had English as their dominant language and had been living in an English-speaking country for the last 10 years or more, we assumed that their parsing preferences would have been comparable to those of the monolinguals in our study.

$^4$ The comparison between Asian collectivistic and non-Asian individualistic participants did not reveal any reliable results in the analyses of gaze direction, fixation delay, and RT in the Sally–Anne task.
(e.g., the word *RIGHT* appeared on the right-hand side). In contrast, in the experimental condition the location of the words was switched (e.g., the word *RIGHT* appeared on the left-hand side). In this second condition, participants had to overcome the impulse to respond according to the location of the word rather than to the word’s meaning.\(^5\)

The task consisted of 50 trials: 10 practice trials and 40 critical trials. All practice trials were in the baseline condition and were not analyzed. Of the 40 critical trials, 20 corresponded with the baseline condition and 20 with the experimental condition. The words appeared in black capital letters (in 54-point Arial font) against a white background on a 15-in. screen. In between trials, the screen remained blank for 600 ms, and then a blue cross appeared in the center of the screen as a fixation point for 1,000 ms. A pseudorandom sequence was generated to present the trials so that the same response would not occur in more than two consecutive trials. One monolingual participant was removed from the data set because of her high error rate (25% of responses; 45% in the critical condition).

**Results.** As in previous studies, bilingual participants showed less interference in the experimental condition relative to the baseline condition than monolinguals (see Figure 5). A 2 × 2 analysis of variance revealed a significant main effect of condition, *F*(1, 43) = 48.68, *p* < .001. The Condition × Group interaction was also significant, *F*(1, 43) = 4.70, *p* < .037, reflecting the superior performance of the bilingual participants in the executive control task. Performance on the baseline condition was comparable for the two groups.\(^6\)

**Correlation Between the Sally–Anne Task and the Simon Task**

For the correlation between the two tasks, we computed fixation duration on the wrong container in the FB condition of the Sally–Anne task. That is, for how long participants fixated on the wrong container from the point when the containers reappeared on the screen at the end of the FB question until they responded to the question. This measure therefore includes first and later fixations. One bilingual participant was eliminated from the analysis because of calibration problems that affected this measure, and one monolingual participant who did not make any eye movements during the processing of the FB question was eliminated. Fixation duration on the wrong container did not reveal a reliable difference between bilinguals (*M* = 206 ms) and monolinguals (*M* = 262 ms), *t*(42) = 0.961, *p* = .342.

We took fixation duration on the wrong container of the FB condition as a measure of egocentric bias, and as such we expected it to positively correlate with participants’ level of executive control. That is, the larger the fixation duration on the wrong container in the FB condition of the Sally–Anne task, the larger the difference between the experimental and the baseline conditions in the Simon task. We calculated the participants’ level of executive control as the difference between their RTs in the baseline and the experimental conditions (*M*\(_{\text{bilinguals}}\) = 26 ms; *M*\(_{\text{monolinguals}}\) = 49 ms). The correlation between the Sally–Anne task and the Simon task was indeed significant for bilinguals, *r*(20) = .513, one-tailed *p* < .008, and for monolinguals, *r*(19) = .529, one-tailed *p* < .008.

Of the 43 participants who were included in the correlation analysis, 12 (seven bilinguals and five monolinguals) did not fixate at all on the wrong container while processing the FB question. Without these 12 participants (for whom fixation duration was 0 ms), the correlation between the Sally–Anne task and the Simon task was still significant for bilinguals, *r*(13) = .761, one-tailed *p* < .001, and for monolinguals, *r*(14) = .491, one-tailed *p* < .028.

**Discussion**

Our overall results show that adults suffer from an egocentric bias in FB reasoning. Unlike in the case of young children, however, this bias does not affect task performance: Adults always answer the FB question correctly. However, with the use of an eye-tracking technique, we observed that the majority of adults momentarily consider the egocentric response before correcting this tendency and taking the perspective of the story character. These results are in line with previous studies revealing an egocentric bias in several different areas of cognition (Birch & Bloom, 2007; Camerer et al., 1989; Keysar, 1994; Keysar et al., 2003).

\(^5\) The version of the Simon task that we used also has elements of the Stroop task, given the conflict between word meaning and word position. We are grateful to Helen Bialystok for pointing out to us that both types of task have shown bilingual advantages.

\(^6\) The comparison between Asian collectivistic and non-Asian individualistic participants in the Simon task did not reveal a reliable difference between the two groups.
Regarding differences between bilinguals and monolinguals, we found that adult bilinguals suffer less from the curse of knowledge than do monolinguals. This advantage in FB reasoning may be maintained from an early age, because bilingual children outperform monolinguals in FB tasks (Goetz, 2003; Kovács, 2009). One of the reasons for this advantage may be the bilinguals’ higher level of executive control, which was confirmed in the present study with the Simon task. Moreover, performance in the executive control task correlated with performance on the Sally–Anne task. An enhanced executive control would help bilingual participants inhibit their own knowledge of the situation in FB tasks, making it easier to take the perspective of the story character (Kovács, 2009).

A second factor that may account for the bilinguals’ advantage in FB reasoning is their early sociolinguistic awareness of their interlocutors’ language background (Comeau, Genesee, & Mendelson, 2007; Genesee et al., 1996). Culture, for example, has been found to have an effect on adult perspective-taking abilities in referential communication tasks, with people from collectivistic cultures suffering less interference from their own visual perspective than do people from individualistic cultures (Wu & Keysar, 2007). If the interdependence that is characteristic of collectivistic cultures can make Chinese university students better at perspective-taking tasks than American students, having to monitor for the language background of their interlocutors from an early age might also make bilinguals better perspective takers than monolinguals.

Future research should determine whether the advantage in FB reasoning that we have observed in adult bilinguals relative to monolinguals is maintained throughout development from early childhood (Goetz, 2003; Kovács, 2009). Bilinguals’ advantage in executive control tasks has indeed been found in children, young adults, and older adults (Bialystok et al., 2004; Carlson & Meltzoff, 2008; Costa et al., 2008). This advantage could in principle help bilinguals maintain an advantage in FB reasoning during development. Moreover, a recent study has shown that the perspective-taking abilities used in referential communication tasks continue to develop into late adolescence in monolinguals (Dumontheil, Apperly, & Blakemore, 2010). It remains an empirical question whether the perspective-taking advantage that bilingual children have shown in language-switch situations, for example (Comeau et al., 2007; Kovács, 2009), would also be maintained throughout development and perhaps enhance their FB reasoning abilities throughout the life span.

Overall, our results support previous research suggesting that bilingualism is one of those life experiences that leads to general cognitive effects (Bialystok, 2009; Bialystok & Craik, 2010). Although these effects have been mixed in other areas of cognition, the effect of bilingualism on FB reasoning is a positive one.

References


Received July 16, 2010
Revision received July 1, 2011
Accepted July 5, 2011