Early false-belief understanding in traditional non-Western societies

H. Clark Barrett, Tanya Broesch, Rose M. Scott, Zijing He, Renée Baillargeon, Di Wu, Matthias Bolz, Joseph Henrich, Peipei Setoh, Jianxin Wang, Stephen Laurence

Electronic Supplementary Material

Table of contents

1. Field sites

Our study was conducted in three different societies: Salar (China), Shuar/Colono (Ecuador), and Yasawan (Fiji), between June 2008 and September 2009. A fourth fieldsite, in Kenya, did not produce useable data (see §6).

As discussed in the main text, studies using spontaneous-response false-belief tasks have altered the previous consensus about the age at which false-belief understanding develops. Prior work using elicited-response tasks led to the view that false-belief understanding is typically not present until approximately age 4 or 5 in the West, and as late as 7 cross-culturally *(13)*. In contrast, spontaneous-response tasks have provided evidence that false-belief understanding is present much earlier, in the third, second, and even first year of life, depending on the task *(19)*. However, all prior work using spontaneous-response tasks has been done in societies that are, according to an acronym coined by Henrich, Heine, and Norenzayan *(34)*, WEIRD: Western, Educated, Industrialized, Rich, and Democratic. As Henrich et al. argue, WEIRD societies probably represent outliers in human culture and human history, differing in many ways from what typical human societies are like in non-industrialized regions of the world, and from what they have been like through most of human history and prehistory. Thus, inferences about human universals – including putative universals of human development, such as the early emergence of false-belief understanding suggested by spontaneous-response studies – cannot reliably be made without cross-cultural evidence, and in particular, evidence from non-WEIRD societies: i.e., traditional, small-scale, non-Western, non-industrialized societies. This was the motivation for conducting our studies at field sites among the Salar, Shuar/Colonos, and Yasawans, all of which differ from the Western settings in which prior spontaneous-response studies have been done in important and theoretically relevant ways.

Developmental psychologists and anthropologists have proposed a variety of ways in which cultural factors – including factors that vary amongst our three study populations and between them and prior study populations in the West – might influence development of mental-state understanding, including understanding of false beliefs. A variety of linguistic factors have been proposed, such as the availability and frequency of use of mental-state terms in the lexicon, including words for "belief" and, in some cases, separate linguistic markers for true and false belief *(38)*, as well as syntactic devices such as complement-taking verbs (e.g., "X thinks that Y") *(39)*. Another is the frequency and nature of conversations between parents and children, including parents' explicit reference to mental states *(40, 41)*, as well as the frequency and nature of joint attention and shared intentionality, as when a child attempts to understand what a parent is referring to *(42)*. Pretend play may also play a role in helping children to entertain the possibility of different points of view on the world, including different beliefs *(43)*. Finally, cultures may vary in a variety of norms and attitudes with respect to minds *(16, 17)*. For example, many cultures in Melanesia and the South Pacific – the culture area including Fiji – are thought to hold norms of "mental opacity" in which individuals refrain from inferring what other people are thinking unless they verbalize their intentions *(18, 44)*. Similarly, among the Gusii of Kenya, LeVine writes that "the Gusii avoid 'psychologizing,' preferring to talk about the overt behavior of adults and children" (pp. 82-83 in *(45)*). Other cultures, including some Mayan cultures of Central America, are said to explicitly disregard mental states and discourage fantasy and pretend play, potentially influencing "theory of mind" development *(46)*. In Samoa, Duranti

(47, 48) has argued that listeners do not primary interpret speakers' utterances in terms of their intentions, which contrasts sharply with proposals that interpretation of speech acts depends heavily on inferences of speakers' intent, at least in some societies *(5)*.

If linguistic and social inputs are important to the development of theory of mind, then cultural variation in linguistic and social practices such as these might lead to different developmental trajectories of false-belief understanding. And there is reason to think that among the world's societies, WEIRD cultures may lie at the far end of the spectrum of explicit, everyday talk and speculation about mental states; many Americans, for example, discuss their feelings, beliefs, desires, and motivations on a daily basis, and regularly engage in discussion about others' mental states as well.

For these reasons, studies of early false-belief understanding in diverse non-Western cultures provide a strong test of the hypothesis that early false-belief understanding, as measured by spontaneous-response false-belief tasks, is a human universal. Given that there are likely to be substantial differences in the nature of linguistic factors, parent-child interaction, and cultural norms both between these societies and between "the West and the rest," similar performance on spontaneous-response tasks at early ages across these societies provides strong evidence for the existence of a universal developmental trajectory of early false-belief understanding, one that is similar across cultures despite significant differences in language, parent-child interactions, and norms regarding others' minds.

While our sample of societies comprises only three cultures and is therefore relatively modest, we selected our sample with an eye towards (1) differences from the typical Western settings in which prior spontaneous-response studies had been conducted, and (2) diversity within our study sites in terms of culture, language, and environment. Here we discuss similarities and differences between our study populations, before proceeding to a more detailed discussion of each site.

Our three field sites were (1) a community of Salar villages in northwest China; (2) a community of villages in the Amazon region of southeastern Ecuador, composed of native Amazonians (Shuar) and "Colonos," or settlers, from the Andean sierra; and (3) a community of Fijian villages on Yasawa Island in northwest Fiji (see details on each site below). These societies all differ from the West along the five criteria captured by Henrich at al.'s WEIRD acronym. Whereas children previously tested on spontaneous-response tasks in the West came from Western, Educated, Industrial, Rich (relatively speaking), Democratic populations, all of the children in our study grew up in families that were non-Western, with relatively low levels of education, not engaged in the industrial economy (all three sites were predominantly engaged in small-scale agriculture, either for personal consumption or sale at local markets), relatively poor, and not engaged in large-scale democracy (Ecuador and Fiji are democratic, but social life at all of our sites is predominantly regulated by community-level interpersonal interactions).

Prior studies of early false-belief understanding in the West have mostly been conducted at universities, with children recruited from local urban populations *(13)*. By and large, these children are exposed to environments with which many readers of this paper will be familiar. Beginning in infancy, Western children are typically surrounded by computers, toys, and extensive exposure to media in the form of television and games. Increasingly, their lives and

education are shaped by parents from an extremely early age, including activities aimed at accelerating cognitive development, early engagement in programs designed to prepare them for formal education, and a surrounding culture that encourages parents to micromanage their children's environments in an effort to improve developmental outcomes. In contrast, children at all of our study sites grew up in traditional, rural, non-Western settings with relatively little wealth, little exposure to Western media, and varying traditional beliefs about child care. In all three of our study populations, traditional attitudes towards child care tend to be more laissez-faire at early stages of development than in the contemporary West. This means, in general, less formally organized parent-child "activities" than in the West, less explicit teaching of children, and thus, less parent-child conversation explicitly intended to instruct the child and guide the child's development *(49)*. If these are factors that are important in shaping early theory of mind development, as has been proposed, then our study sites are well-situated to detect differences in early understanding of false beliefs. Moreover, in all of our study sites, children tend to be treated as having less personal "agency" than in the West: at early ages, children are accorded less personal choice and autonomy than children in the West might be. This means that there may be less conversation between parents and children about their preferences and desires, also suggested to play a role in theory of mind development *(41)*. Finally, children at our three field sites probably interact with a smaller number and diversity of social partners, including strangers, meaning less exposure to divergent points of view, another possible factor in theory of mind development *(43)*.

In addition to these commonalities, which set our study populations apart from the Western settings where prior work on early false-belief understanding has been conducted, our three populations also differed in many ways. Geographically, and environmentally, of course, they are quite distant: the Salar live in northwest China, the Shuar/Colonos live in the Amazon, and the Yasawans live on a small island in the South Pacific. In terms of cultural and linguistic phylogeny – i.e., relationships of descent from a common ancestral culture – one can hardly imagine three more distantly related cultures and languages. While we have not directly measured some aspects of language and linguistic norms hypothesized to influence development of belief understanding, such as the prevalence of complement-taking verbs or the prevalence and diversity of mental-state terms, there is no question that the languages themselves are distantly related: Salar, a Turkic language influenced by Tibetan and Chinese; Spanish, a European Romance language (in which the studies were conducted at our Shuar / Colono site – many participants were also fluent in Shuar, a member of the Jivaroan language group); and the Teci dialect of Fijian, a member of the Malayo-Polynesian language family. Moreover, many cultural practices among these three groups are quite different, ranging from subsistence practices to social practices such as family structure and marriage rules. While all three cultures probably engage in less direct pedagogy of children than Western cultures do, and all regard children as having less agency and self-awareness than do Western cultures, the cultures differ in their childcare practices. What is expected of children (e.g., household chores and obligations, expectations about how to relate to other children and adults) varies across the cultures, along with norms about how to properly raise children. For example, the Salar believe it is not good for children to receive too much care from grandparents, a belief not present among Shuar or Yasawans.

Attitudes about the mind, and the degree and nature of talk about the mind – held by

anthropologists and psychologists alike to be potentially important in influencing development of understanding of false beliefs *(16, 50)* – vary across our study sites. The Salar value frankness and frequently speak about internal states. The Shuar, on the other hand, are more reserved in both talking about their own internal states and speculating about the mental states of others. Yasawans are in the culture area in which norms of mental opacity are thought to be common – i.e., it is held that it is not possible to know what is inside the minds of others *(18)* – and while adults do not appear unwilling to discuss internal states, they rarely talk about the mental states of children, and indeed regard very young children as incapable of thought or feeling. These and other aspects of each of our three study populations are discussed in more detail below.

1.1 Salar. The Salar portion of the study was conducted in Xunhua Salar Autonomous County, Qinghai Province, China (Figure S1). The Salar, whose total population is about 100,000, live mainly in this county, and they are one of 55 minority ethnic nationalities in China. Our participants were sampled from two villages, Yangkulang and Gujilai, which are clusters of village houses belonging to about 370 Salar households with a total population of about 2,100 people.

Figure S1. Salar field site, China.

Traditionally, the economic base of Salar villagers was agriculture, and they have made ends

meet by producing wheat, cultivating fruit trees, and raising domestic animals. However, from 1990, the rapid increase of the local population created an imbalance between the population and farmlands. As a result, the income from agricultural production became limited, causing villagers to gradually quit farming and instead, to leave the villages to search for a new way to make a living. Some started businesses relating to agricultural production, and many people became wealthy by raising and selling livestock. Now many villagers run businesses raising, selling, and transporting domestic animals, or run businesses dealing with real estate and transportation. These developments have greatly benefited the economy of the villages. The primary language of the Salar is an eastern dialect of the Turkic language group, part of the Altaic Language family. The Salar use Chinese characters in writing. Although the Salar receive public education in mandarin Chinese, many young children and uneducated adults have a relatively poor understanding of mandarin Chinese. Therefore, the current study was conducted in the Salar dialect. Each village has its own primary school, with about 600 pupils enrolled from the first to the sixth grade as regular students, accounting for more than 95% of the total children in the villages. However, education beyond primary school is much more sporadic, since many villagers send their children away from the villages after primary school. There are perhaps several dozen junior and senior high school graduates as well as some university graduates among the villagers.

In many ways, the social and physical environments of children growing up in the two Salar villages are similar to those of many children growing up in small, rural villages in other non-industrialized parts of the world. Traditionally, young children spend lots of time with their parents. In a typical Salar family, an older brother will remain at home with his parents after marriage, but will move out from the parents' home and start his own independent household with his wife and children, as soon as a younger brother gets married. Moreover, many Salar parents believe that children who are brought up by grandparents are often headstrong, because grandparents are usually too soft-hearted to discipline them. However, recently more and more young parents have migrated to work in big cities. Those parents leave their children with paternal grandparents. Therefore, nowadays some Salar children also interact with grandparents on a daily basis. Overall, children rarely interact with strangers outside of their family or their village, and children probably interact with a smaller number of daily social interactants than is typical of more urban settings. Being in a rural setting, young children's activities change based on the season. In autumn, children enjoy spinning tops, and in winter, they play a game with stones on the ice. Some urban toys and activities have also been introduced to the Salar villages, such as flash cards and yo-yos. In recent years, many children have begun to have some access to technologies such as internet and DVDs.

For the Salar, some aspects of the social environment that they experience in early childhood may differ from other cultures. Unlike Westerners, the Salar parents do not encourage young children's independence. The Salar parents make decisions and even directly solve problems for their children until they reach adulthood. Children are typically not asked their views or opinions for decision making. In addition, having a nomadic ancestry, the Salar people value frankness. Compared to many Westerners, the Salar people tend to talk more openly about their internal states, such as their intentions, emotions, preferences and so on. The Salar parents have a relatively more laissez-faire attitude towards infants' behavior than do typical American parents. According to their religious beliefs, boys younger than 12 years and girls younger than 9 years

are not expected to fulfill any religious or social obligation. Therefore, parents are not strict with young children's exploratory behavior. For these reasons, the Salar may be an interesting population for the study of early theory of mind development.

We conducted our study with the assistance of a local family, in the living room of their home, a single-story brick house. All of the testing was conducted by local research assistants, in the Salar dialect. Two researchers (Z.H. and D.W.) supervised the testing and operated the video recording equipment.

1.2 Shuar/Colono. Our Shuar/Colono participants were recruited from villages in Cantón Palora, Morona Santiago Province, Ecuador (Figure S2). This region of Ecuador is known as the "Oriente," the Eastern, or Amazon region. The Cantón (County) of Palora is approximately 1,500 km^2, with a little over 6,000 inhabitants at the time of the 2001 census, and an average elevation of around 900 m. The local population consists of three main ethnicities: "Colonos" (colonists) from the Sierra (of mixed European and indigenous descent), and two indigenous groups, Quechua and Shuar. The majority of our participants were Shuar, with a smaller number of Colono participants, and a few children of Quechua descent.

Figure S2. Shuar / Colono field site, Ecuador.

Traditionally, the Shuar are a hunter-horticulturalist population, living in small villages ranging in size from a single extended family to a collection of perhaps twenty or so households *(51)*. Until the late 1960s, the Shuar of the Palora area lived in the traditional way, hunting for forest game with blowguns, fishing with nets and plant piscicides, and maintaining household gardens that supplied manioc, plantains, and taro, the main caloric staples of the Shuar diet. In 1967, the town of Palora was established by Ecuadorian Colonos from the Sierra (Andes), encouraged to migrate into the area. With the arrival of the Colonos, life slowly began to change as roads were established between Palora (today a town of several thousand) and the nearby Shuar communities. While most Shuar people continued to live in largely traditional fashion in small villages, they gradually became more dependent on the market economy, and today many Shuar maintain small plots of cash crops, such as sugar cane, for sale in the Palora market. In the Shuar villages of the Palora region, the predominant language is still Shuar, though all Shuar people in the area are bilingual (i.e., they also speak Spanish), thanks to a government-sponsored bilingual education program. Most young children remain in their village until at least their teenage years, and perhaps through adulthood. Until recently, most Shuar did not have beyond a first or second grade education, though now it is becoming more common to continue education up to, and sometimes through, high school. This can be a main reason for leaving one's village – some children go to Palora for high school, staying with relatives. In addition, many Shuar families now live and work in Palora itself.

In many ways, the social and physical environments of children growing up in this area are similar to those of many children growing up in small, rural villages in other non-industrialized parts of the world. Life is primarily based around families, and young children spend lots of time with their mothers as they conduct their daily household routines, as well as mixed-age groups of older siblings *(52)*. The majority of social interactions, then, are with familiar faces, and children probably interact with a smaller number of daily social interactants than is typical of more urban settings. Being in a rural setting, young children's activities include lots of observation of subsistence activities like gardening, cooking, and animal husbandry, as well as indoor and outdoor play, mostly with siblings. There is little or no contact with electronic media, save for some observation of television in some homes (there are now a few television sets even in the smallest village, powered by generators and used for watching DVDs when off the grid), but almost no contact (at the time of the study) with the internet, videogames, or smartphones. The above observations hold true for most of the children in our sample, with the exception of some children from more well-to-do families in the Palora area, who have greater access to technological toys like video games.

For the Shuar, some aspects of the social environment in early childhood may differ from other cultures. Like many traditional cultures, the Shuar appear to engage in relatively less speech directed specifically towards young children than, for example, urban Americans might. Young children are typically not asked their views or opinions about things, such as whether or not they like their food, or if they know how some artifact works, as American children might be. Also, compared to many Westerners, who eagerly report their moods and other internal states, the Shuar tend to make fewer reports about their moods and feelings. As in some other Amazonian societies, speculation about states of affairs of which one does not have direct knowledge is frowned upon (speculating about things one hasn't observed, for example, is sometimes considered "lying," and there is a strong norm against lying) *(53)*. This includes speculating

publicly about the intentions and other internal states of others. This might lead to less exposure to mental-state terms overheard in others' speech. Moreover, perhaps because of the constraints of a difficult-to-manage environment, Shuar parents typically have a much more laissez-faire attitude towards infants' exploratory behavior than do typical American parents; for example, as in many traditional cultures, children are allowed to handle dangerous artifacts such as machetes or knives at a much earlier age (e.g., 4 or 5 years old) than most Western parents would allow (though note that Shuar parents keep a close and watchful eye on infants to prevent injury). There may thus be less one-on-one interaction between parents and children in children's play, and therefore fewer episodes of "shared intentionality," suggested by some authors to play an important role in the development of theory of mind *(42)*. For these reasons, the Shuar are an interesting population for the study of early theory of mind development. While the above cultural considerations do not hold to the same degree for the non-Shuar Ecuadorian children in our sample, it is nevertheless the case that the non-Shuar children share many aspects of their physical and social environments with the Shuar, including relatively small social communities, modest levels of education (high school education only now becoming common) and low income levels (typical income at the time of the study was around $10/day or less).

We conducted our study with the assistance of the local, government-run daycare system. Daycares in this area vary in quality, but they are typically a two- or three-room house made either of wood or cinder block, with a zinc roof, staffed by local mothers. They are usually located in the village itself, in close proximity to the houses where the children live. Children in these daycares typically engage in group play, under the supervision of the daycare staff, until around mid-day, allowing their parents to work and do chores. We conducted our study in 9 daycares in the Cantón Palora: four were in Shuar villages, three were in small mixed-ethnicity settlements, and two were in the town of Palora itself. Because our study was conducted with a mix of Shuar and non-Shuar children, all of the testing was performed in Spanish, by Spanish-speaking American research assistants. Two additional researchers (H.C.B. and M.B.) supervised the testing and operated the video recording equipment.

1.3 Yasawan. The Yasawan portion of the study was conducted in three neighboring villages in the Yasawa Island Group of the Fijian islands (Figure S3). The Yasawa Island Group is made up of five island clusters with Yasawa Island being the farthest island in the archipelago. There are 6 villages on the island, with between 100 and 250 people per village. Whereas much of the Yasawa Island Group relies on tourism, travel to this particular island is difficult and costly therefore visits from outsiders are rare. There is minimal Western cultural influence with no newspapers or electricity in these villages at the time of this study. Villagers in this region live in traditional thatched houses and rely on subsistence agriculture and marine foraging and fishing for livelihood. Adult daily life consists of maintaining gardens, harvesting, fishing and meal preparation and cooking on an open fire. Food preparation takes up a significant proportion of time and children are often expected to accompany adults and observe or participate quietly during these activities.

Figure S3. Yasawan field site, Fiji.

Yasawan children are typically regarded as the lowest ranking members of Yasawan society. Not only are children regarded as low ranking, parents also report that children do not 'understand' language and cannot 'think' or 'feel' pain or pleasure until well into their second year of life *(54)*. This contrasts with American parental reports indicating that infants are able to think and feel pain or pleasure at birth (or before). This difference in attitudes about developmental milestones has been demonstrated in many non-Western/Western comparisons of parental beliefs about development *(55)*. Yasawan parents' beliefs are also reflected in their behavior, as they engage in very little face-to-face dialogue with infants and young children in day to day life (as with many other rural or small-scale societies; *54)*. It is also uncommon to have discussions about the feelings or thoughts of a child. Children in many south pacific islands are expected to be seen but not heard, and this is also the general way of Fijian parenting styles and family life *(56)*. In general, child life is similar to that in many rural, small-scale cultures, with children engaging in outdoor play with limited adult supervision and accompanying adults during their daily domestic activities.

The verbal preferential-looking and anticipatory-looking tasks were given to children in three

villages (Teci, Yasawa i Ra Ra, Bukama) on Yasawa Island; each village had a population between 100-250 people at the time of the study. All three villages are within walking distance (although it takes a day to walk from Teci to Yasawa i Ra Ra). In each village, a temporary laboratory was established in a local home. All of the testing was conducted by local research assistants, in the language of the region. Although Fijian (also referred to as Bauan) is spoken by inhabitants of these islands, infants and young children (before formal education) are more familiar with the local dialect which is specific to each village. As the villages are near one another, the local dialects are very similar. We tested all children with our 'Teci' script (the local dialect language spoken in the Teci village), but we modified it for Yasawa i Ra Ra and Bukama participants, based on consultation with the mothers regarding the languages that their children were most familiar with. In short, we tested all children in a local dialect of Fijian, specific to each village. One additional researcher (T.B.) supervised the testing and operated the video recording equipment.

For the non-verbal violation-of-expectation task, children were tested on the outskirts of Lautoka, a small city on the mainland of Fiji; the children were tested in standard Fijian (Bauan). As explained below, the data from that task were not usable (trials were ended too soon, due to experimenter error).

2. Materials and Methods: General Remarks

We adapted three spontaneous-response false-belief tasks that have produced positive results with young children tested in Renée Baillargeon's Infant Cognition Laboratory at the University of Illinois: a verbal preferential-looking task *(22)*, a verbal anticipatory-looking task *(21)*, and a largely non-verbal violation-of-expectation task *(25)*.

In adapting these tasks for our study, we made two important changes. First, for each task, we chose a somewhat broader age range than had been used in Illinois, to maximize the number of children that could be tested at each field site. We were uncertain how this decision would affect our results: in any given task, children who were younger or older than those tested in Illinois might fail, not because they lacked false-belief understanding, but because the procedure used was ill-suited for their particular age group (e.g., provided too little familiarization with the task for the younger children, or progressed at too slow a pace for the older children). Nevertheless, given the difficulties of collecting data at our various field sites, it made sense to collect more data rather than less.

Second, in each task, we decided to test children only in the critical false-belief condition. The three tasks used in Illinois also included a true-belief condition in which the agent held a true, instead of a false, belief about the scene. In all tasks, reliable differences were obtained between the false- and true-belief conditions: children expected the agent to act on her mistaken belief in the false-belief condition (e.g., to search for the object where she falsely believed it was still hidden), but to act on her knowledge in the true-belief condition (e.g., to search for the object where she knew it was currently hidden). Together, the true- and false-belief conditions provided evidence that children expected the agent to act on whatever information was available to her, whether this information was in fact correct or incorrect. Because we were primarily interested in assessing early false-belief understanding in our different populations, all children were tested in

the false-belief conditions of the tasks. Since experiments of this type had never been conducted on young children in field conditions, at the outset of the study we were uncertain how many children could be tested at each field site, how much of the data collected would be usable given imperfect testing conditions (e.g., noisy, distracting surroundings), and how suitable our procedures would be for the broad age ranges selected; collecting only false-belief condition data thus maximized our ability to assess false-belief understanding in each population. We reasoned that if results comparable to those from Illinois were obtained in each task and at each site, it would provide robust evidence that Salar, Shuar/Colono, Yasawan, and Illinois children possessed a similar capacity for early false-belief understanding, as assessed by spontaneous-response tasks.

For each task, we begin by describing the procedure used in Illinois; we then describe how the procedure was adapted to each field site.

3. Verbal Preferential- Looking False-Belief Task

3.1 Participants

In the Illinois sample *(22)*, participants in the verbal preferential-looking false-belief task were 29- to 34-month-olds ($M = 31.6$ months). To maximize the number of young children that could be tested at each field site, a broader age range was used that included children between 25.5 and 52 months.

Salar participants were 25 children, 14 male and 11 female, ranging in age from 29.3 to 51.1 months ($M = 40.0$ months). Another 7 children were tested but excluded, 1 because of parental interference, 1 because of failure to complete the task, 1 because of side bias (spending more than 85% of the testing session looking to the same side), and 4 because the difference in their looking times at the two test pictures was over 3 standard deviations from the mean. (It is unclear why there were 4 outliers in this task among the Salar, as there were few outliers overall in other tasks or at other field sites. Given that 3 of these 4 outliers were among the last 4 children tested, we suspect that procedural variations that could not be detected on the videos might have affected the children's responses).

Shuar/Colono participants included both a younger group in the same age range as the children at the other two sites, and an older group. Participants in the younger group were 58 children, 26 male and 32 female, ranging in age from 25.8 to 51.8 months ($M = 39.8$ months). Another 7 younger children were tested but excluded, 4 because of side bias, 2 because of inattentiveness (spending less than 30% of the testing session looking at the pictures), and 1 because of a possible developmental delay. Participants in the older group were 20 children, 14 male and 6 female, ranging in age from 52.2 to 64 months ($M = 58.2$ months). One additional older child was tested but excluded due to side bias.

Yasawan participants were 11 children, 4 male and 7 female, ranging in age from 25.8 to 43.4 months ($M = 32.5$ months). Another 2 children were tested but excluded, 1 because of inattentiveness and 1 because the difference in his looking times at the two test pictures was over 3 standard deviations from the mean.

3.2 Procedure in Illinois

This task took advantage of children's well-established tendency to look preferentially at scenes that match the utterances they hear. Children sat on a parent's lap at a table in a testing room, with a picture book raised at an angle in front of them. An experimenter sat to the child's left. A camera across from the children captured the children's eye movements. The picture book consisted of eight double-pages attached to the top of a bookstand with three binder rings. Each double-page (55 cm x 22.5 cm) was composed of two clear plastic photo sheets attached in the center; color photos (20 cm x 25 cm) were inserted in the sheets and were spaced about 5 cm apart.

At the start of the experiment, the double-pages were face down behind the bookstand. In each trial, the experimenter recited a line of the story, flipped the page so that the photos became visible to the children, and then repeated the line of the story. The experimenter then paused for approximately three seconds before reciting the next line of the story. Throughout the story, the experimenter gazed at a neutral location so that the children could not use her gaze as a cue for finding the relevant picture.

The story was organized into three parts: introduction, setup, and test. The introduction trials presented the main character, Emily (introduction-1), and her friend, Sarah (introduction-2). In these two trials, one side of the double-page displayed a picture and the other displayed white paper. In the subsequent trials, one side of the double-page displayed a picture that matched the story (matching picture) while the other displayed an irrelevant picture (non-matching picture). In the setup trials, children were told that Emily had an apple (setup-1) and that she hid it in one of two containers (setup-2). Emily then went to take a nap (setup-3). While Emily was gone, Sarah moved the apple to the other container (setup-4) and then went outside to play (setup-5). The story then ended with a test trial in which Emily returned to look for her apple. One picture showed Emily searching for her apple in its initial location, and the other showed Emily searching for the apple in its current location. Hiding container, side of matching picture in the setup trials, and side of the initial-location picture in the test trials was counterbalanced across children. For full script, pictures, and additional details of the procedure, as well as descriptions of control conditions, see *(22)*.

3.3 Adaptation to field sites

At each site, the children sat at a table either next to or on the lap of a caretaker who was known to them. The script was adapted to use culturally appropriate character names, activities, and objects. The photos at each site were taken with locally recruited actors who were unknown to the children. At each site, the side of the initial-location picture in the test trial was counterbalanced across children. Below, we list additional modifications at the different sites.

Salar. In addition to the shared site variations, the procedure used with the Salar children differed from that in Illinois in three ways. First, the experimenter sat on the children's right rather than on their left. Second, the photos were slightly larger (21 cm x 28 cm), and they were spaced farther apart (16 cm) to facilitate coding. Third, the introduction trials involved four

single-picture trials rather than two. In each added trial, the photos depicted one of the main characters doing a common household activity. This change was made to give children additional time to adjust to the unusual storybook and the fact that the relevant photo might appear on either side. Side of matching picture in the introduction and setup trials was counterbalanced across trials and across children. In addition, hiding container (basket or bucket) was also counterbalanced across children at this field site.

Shuar/Colono. In addition to the shared site variations, the procedure used with the Shuar/Colono participants differed from that in Illinois in three ways. First, the experimenter sat across from the children rather than to their left (the experimenter never made eye contact with the children). Second, the photos (21 cm x 25 cm) were spaced farther apart (30 cm). Third, the book was set up very differently at this field site. In Illinois, and with the Salar and Yasawan populations, the pictures were attached to the bookstand at the top and flipped forward towards the children. This meant that both sides of the double-page became visible to the children at the same time. For the Shuar/Colonos, the picture book was designed like an actual book and the pages turned from right to left. This meant that as the experimenter turned the page, the picture on the children's right became visible before the picture on the children's left. Because the coding window began after the page was fully turned (see Coding section), this meant that children were able to inspect the right picture before the coding window began. In addition, the experimenter had to hold the left pages so that the book remained open and his hand thus covered a small portion of the photo on the children's left. These differences had a significant impact on children's performance in the test trial. At the start of the trial, children had a tendency to look longer at the picture on their left, regardless of whether that was the initial- or current-location picture. This was likely due to some combination of the fact that (1) this picture was visible second and children had a tendency to look at the pictures in the order that they became visible and (2) the experimenter's hand touched that picture and captured children's attention briefly. However, closer inspection of the data revealed that after several seconds, children overcame this bias and switched to looking longer at the initial-location picture, regardless of whether it was on the left or the right. We therefore divided the trial into two four-second windows. The first window served as a preview window, allowing children to inspect the pictures and overcome the side bias induced by the procedure. The second window served as the test window used in the data analyses.

Yasawan. In addition to the shared site variations, the procedure used with the Yasawan children differed from that in Illinois in four ways. First, the photos were slightly larger (21 cm x 29.7 cm) and spaced farther apart (about 10 cm). Second, the setup trials involved two additional two-picture trials. In one trial, the first story character was eating his breakfast; in the second, the second story character was cooking. As with the Salar, these additional trials were added to give the children additional time to adjust to the storybook and the fact that the relevant photo might appear on either side. Third, rather than memorizing the story, the experimenter read the story from a script on a piece of paper that was visible to the child. Fourth, the experimenter told the story much more slowly than was done in Illinois, pausing for approximately 12 seconds (instead of for three seconds) before each new line of the story.

Figure S4 shows the stimulus materials for one site (Shuar/Colono).

introduction trials	*Esta es la história de una señorita que se llama Noemí. Ahí está Noemí. Vé? Ahi esta Noemí.* This is the story of a girl named Noemi. There's Noemi. See? There's Noemi.	
	Noemí tiene una amiga que se llama Jaqueline. Ahi está Jaqueline. Vé? Ahí está Jaqueline. Noemi has a friend named Jaqueline. There's Jaqueline. See? There's Jaqueline.	
setup trials	*Noemí tiene una naranja. Vé? Noemí tiene una naranja.* Noemi has an orange. See? Noemi has an orange.	
	Noemí pone su naranja en un barro. Vé? Esta poniendo su naranja en el barro. Noemi puts her orange in a vase. See? She's putting her orange in the vase.	
	Noemí está cansada. Se acuesta en la cama para dormir. Vé? Está durmiendo. Noemi is tired. She lies down in the bed to go to sleep. See? She's sleeping.	
	Cuando Noemí está durmiendo, Jaqueline saca la naranja del barro y la pone en la olla. Vé? Está poniendo la naranja en la olla. While Noemi is sleeping, Jaqueline takes the orange out of the vase and puts it in the pot. See? She's putting the orange in the pot.	
	Jaqueline sale de la casa para alimentar a los pollos. Vé? Jaqueline está alimentando a los pollos. Jaqueline leaves the house to feed the chickens. See? Jaqueline is feeding the chickens.	
test trial	*Noemí se despierta y tiene hambre. Busca su naranja. Vé? Está buscando su naranja. (Pausa 3 segundas). Está buscando su naranja.* Noemi wakes up and is hungry. She looks for her orange. See? She is looking for her orange (Pause 3 seconds). She is looking for her orange.	

Figure S4. Verbal Preferential-Looking False-Belief Task Booklet — Shuar/Colono version. The two columns on the right show pictures that were simultaneously presented (each row shows a pair of simultaneously presented pictures in the story book; in the introduction trials, only one photo was shown at a time). Text to the left of the photos was read aloud to the child.

Figure S5. Verbal Preferential-Looking False-Belief Task being run at Yasawan field site.

3.4 Coding

In each two-picture trial, we coded where children looked (left picture, right picture, away) frame-by-frame from silent video. For the setup trials, we coded the first eight seconds that the pictures were visible to the children. This window ended during the pause after the story line was repeated, prior to the story line for the next trial. In the test trial, we coded either the first four seconds that the pictures were visible to the children (Salar, Yasawan), or the first four seconds following the preview window (Shuar/Colono), as explained above.

All test trials were coded independently by a second, naive coder. To ensure that this second coder was blind to the participant's condition, all videos were edited to start at the onset of the test trial. Thus, the second coder had no information about where the object was hidden or which picture was on each side in the test trial. Reliability was calculated as the percentage of frames in which the primary and secondary coders agreed on the child's direction of gaze (right, left, or away). In this and all other tasks in the study, trials in which agreement was less than 50% were resolved through discussion (such trials were rare). Reliability was high across all three sites: Salar, 93%; Shuar/Colono, 97%; and Yasawan, 87%. Averaged across sites, the two coders agreed on the children's direction of gaze for 95% of the coded video frames.

Preliminary analyses of the setup and test data revealed no interaction of picture with side condition in setup (Salar only) or hiding container (Salar only). The data were therefore collapsed across these factors in subsequent analyses.

3.5 Results

Setup trials. Figure 1 (see main paper) shows the mean looking times to the matching (story relevant) and non-matching (story irrelevant) pictures averaged across all setup trials, separately by field site. Children's responses were analyzed using paired t-tests (all tests in this ESM are two-tailed). In each population, children looked significantly longer at the matching than at the non-matching picture across setup trials: Salar $t(24) = 9.63$, $p < .001$ (matching $M = 4.33$, $SD = 0.84$; non-matching $M = 2.30$, $SD = 0.60$); Shuar/Colono-younger $t(57) = 16.06$, $p < .001$ (matching $M = 4.64$, $SD = 0.83$; non-matching $M = 1.95$, $SD = 0.63$); Shuar/Colono-older, $t(19) = 9.82$, $p < .001$ (matching $M = 4.81$, $SD = 0.81$; non-matching $M = 1.97$, $SD = 0.65$); Yasawan $t(10) = 4.67$, $p = .001$ (matching $M = 3.82$, $SD = 0.89$; non-matching $M = 2.36$, $SD = 0.65$).

To confirm these effects, we next analyzed the data from each population with an analysis of variance (ANOVA) with picture (matching or non-matching) as a within-subject factor and other variables as between-subjects factors. For the Salar, an ANOVA with sex as a between-subjects factor confirmed the main effect of picture, $F(1, 23) = 101.82$, $p < .001$. No other effects were significant, all Fs < 1.4. For the Shuar/Colonos, an ANOVA with sex and age group (younger, older) as between-subjects factors confirmed the main effect of picture $F(1, 74) = 237.36$, $p < .001$. No other effects were significant, all Fs < 2.3, all ps > .13. Finally, for the Yasawans, an ANOVA with sex as a between-subjects factor again revealed a main effect of picture, $F(1, 9) = 28.11$, $p < .001$. This analysis also revealed a significant picture by sex interaction, $F(1, 9) = 9.77$, $p = .012$. This resulted from the fact that females (matching $M = 4.12$; non-matching: $M = 2.11$) demonstrated a slightly stronger preference for the matching picture than did the males (matching: $M = 3.30$; non-matching: $M = 2.78$). .

Finally, we performed an omnibus linear regression on all our samples pooled, with difference in looking-time between matching and non-matching picture as the dependent variable and age (in months), sex, and sample population (dummy coded, with the younger and older Shuar/Colono children treated as separate samples) as independent variables. This found no significant effects of age, sex, or population within the entire sample pooled, with the exception of the Shuar/Colonos, who had slightly higher overall looking times than the Salar and Yasawans (Shuar/Colono-younger, $\beta = .24$, $p = .04$; Shuar/Colono-older, $\beta = .274$, $p = .07$).

Test trial. Figure 1 (see main paper) shows the mean looking times to the initial-location and current-location pictures during the 4-s test window, separately by field site. Children's responses were analyzed using paired t-tests. Children in all three societies looked reliably longer at the initial-location picture: Salar $t(24) = 2.59$, $p = .016$ (initial-location $M = 2.27$, $SD = 1.07$; current-location M 1.32, $SD = 0.96$); Shuar/Colono-younger $t(57) = 3.16$, $p = .003$ (initial-location $M = 1.94$, $SD = 1.18$; current-location $M = 1.09$, $SD = 1.09$); Shuar/Colono-older, $t(19) = 2.36$, $p = .029$ (initial-location $M = 2.20$, $SD = 1.01$; current-location $M = 1.17$, $SD = 1.12$); and Yasawan $t(10) = 2.26$, $p = .047$ (initial-location $M = 2.36$, $SD = 0.83$; current-location $M = 1.26$, $SD = 0.94$).

To confirm these effects, we next analyzed the data from each population using an ANOVA with picture (initial- or current-location) as a within-subject factor and other variables as between-subjects factors. For the Salar, an ANOVA with sex as a between-subjects factor confirmed the

main effect of picture, $F(1, 23) = 6.44$, $p = .018$. No other effects were significant, all Fs < 1. For the Shuar/Colonos, an ANOVA with sex and age group as between-subjects factors confirmed the main effect of picture, $F(1, 74) = 10.01$, $p = .002$. There was also a marginal effect of age group, $F(1, 74) = 2.80$, $p = .098$, caused by the fact that the older age group looked slightly longer ($M = 1.72$) than the younger age group ($M = 1.51$). No other effects were significant, all Fs < 1.93, all ps > .16. Finally, for the Yasawans, an ANOVA with sex as a between-subjects factor again revealed a main effect of picture, $F(1, 9) = 8.27$, $p = .018$. No other effects were significant, all Fs < 3.1, all ps > .11.

Finally, we performed an omnibus linear regression on all our samples pooled, with difference in looking-time between the initial- and current-location picture as the dependent variable and age (in months), sex, side of initial-location picture, and sample population (dummy coded) as independent variables. This analysis found no significant effects of these variables within the entire sample pooled. Across the four populations, 68% (77/114) of the children looked longer at the initial- than at the current-location picture during the test window, a percentage comparable to the 75% (21/28) obtained by Scott et al. *(22)* in Illinois.

4. Verbal Anticipatory-Looking False-Belief Task

4.1 Participants

In the Illinois sample *(21),* participants in the verbal anticipatory-looking false-belief task were 29- to 34-month-olds ($M = 30.8$ months). To maximize the number of children that could be tested at each field site, a broader age range was used that included children between 29 and 52 months.

Salar participants were 13 children, 7 male and 6 female, ranging in age from 29.3 to 51.1 months ($M = 39.0$ months).

Shuar/Colono participants were 29 children, 13 male and 16 female, ranging in age from 29.1 months to 51.8 months ($M = 40.0$ months). Another 13 children were tested but excluded, 12 because they were uncooperative (e.g., did not respond to the experimenters) or distracted (e.g., played with the testing table or attended to loud noises outside the testing room), and 1 because his target advantage score (looking time at target minus non-target container; see below) was over 3 standard deviations from the mean.

Yasawan participants were 7 children, 3 male and 4 female, ranging in age from 29.9 months to 43.4 months ($M = 36.4$ months). Another 4 children were tested but excluded because they were distracted (e.g., played with their sticker or attended to noises outside the testing room).

Because several Shuar/Colono and Yasawan participants in the verbal anticipatory-looking false-belief task were flagged as uncooperative or distracted by the research assistants at the field sites and by the primary coder who coded the videos, all videos from these two field sites were independently evaluated by two judges. The judges rated each child on an "on-task" scale from 1 (mainly on-task: cooperative child, no notable distraction) to 4 (mainly off-task: very shy or uncooperative child, and/or one major distraction or multiple minor distractions). A child was

excluded from the data analyses if both judges agreed that the child was mainly off-task (4); disagreements between the two judges were resolved though discussion.

4.2 Procedure in Illinois

The procedure used in Illinois included a warm-up phase, a sticker-game phase, a belief-induction phase, and a test phase.

Warm-up phase**.** At the start of the warm-up phase, the first experimenter, E1, stood across from the child; the two containers rested on the table, with the scissors hidden in one of them (target container side was counterbalanced). E1 introduced herself and then used a few warm-up toys to establish rapport with the child.

Sticker-game phase**.** During the sticker-game phase, E1 announced that they would now play a game in which she would give the child stickers and the child would place them on a sheet of paper to take home. As the game developed, E1 removed and replaced props from a bin at her feet. To start, E1 presented the child with several sheets of colored paper, asked the child to choose one, wrote the child's name on the paper with a marker, and reminded the child of E1's own name (it was important for the test phase that the child know E1's name). Next, she presented the child with several sheets of stickers and asked the child to choose a sheet. E1 then said, "I'm going to need my scissors to cut out these stickers for you!" E1 first opened the non-target container, showed the child that it was empty, closed it, and replaced it on the table. Next, she opened the target container where she found her scissors, exclaiming "That's right! This is where my scissors are. This is where I keep them!" After closing the container, E1 asked the child to choose a sticker; she then cut it out and helped the child affix it to the paper. The child then chose a second sticker; as E1 was about to cut it out, she coughed mildly; this served as a signal to the second experimenter, E2 (who was waiting outside the room) to knock at the door. E1 said, "Come in!", and E2 opened the door and told E1 she had a phone call. E1 replied, "No problem, I will be right there!" E2 then closed the door and waited outside. E1 put the scissors away in the target container, put the sheet of stickers back into the bin, and informed the child that she would continue cutting out the stickers when she returned. E1 then opened the door, ending the sticker-game phase.

Belief-induction phase**.** In the belief-induction phase, E1 asked E2 at the door (in the child's hearing) to stay with the child while she was gone. E2 answered, "Sure, no problem!" E2 then walked into the room, closed the door, and greeted the child as she took E1's place at the table. To establish rapport, E2 spent a few minutes chatting with the child (e.g., about the child's sticker). When E2 felt that the child was comfortable, she asked casually, "What else are you playing with? What's in these containers?" She first opened the non-target container, showed it to the child while saying, "Oh, there is nothing in this one!", and closed it again. Next, she reached for the target container while saying, "What about this one? Is there something in here?" She then opened the target container, took out the scissors, and closed the lid while saying, "Oh, scissors, I was looking for a pair just like these! I need them for a project I am working on. I am going to take them!" She then put the scissors in her pocket, ending the belief-induction phase. (In pilot work in Illinois, He and her collaborators found that having E2 take possession of the scissors, as opposed to moving them to the other container, was a salient manipulation for young

children, perhaps due in part to self-interest: if E2 left with the scissors, then E1 could not continue cutting out stickers for them when she returned).

Test phase. In the test phase, E2 looked at the ceiling, chin in hand, and said, as though thinking out loud, "But when [E1's name] comes back, she's going to need her scissors again..." E2 paused for 2 s and then delivered the anticipatory prompt, as if wondering aloud to herself, "Where will she think they are?" E2 then paused for 5 s, in the same position (staring at the ceiling, chin in hand). Following this 5-s *response period*, the testing session ended. For additional details on the procedure and descriptions of control conditions, see *(21)*.

4.3 Adaptation to field sites

At each site, the children sat at a table either next to or on the lap of a caretaker who was known to them. The script was adapted to use culturally appropriate character names and objects. In addition, rather than always telling E1 she had a phone call, E2 used various excuses across sites to say she was needed elsewhere (e.g., someone wanted to talk to her). At each site, the side of the target container was counterbalanced across children.

Pilot data were collected with 2.5 to 6-year-old Shuar children using a response period similar to that in Illinois. Examination of these data revealed that, although the children aged 5 years and older generally performed well (these data are described at the end of the Results section, §4.5), the children under age 5 did not. We speculated that these younger children might require additional time to process the anticipatory prompt, perhaps because it was often delivered at a faster rate than in Illinois, or perhaps because of some other factor (e.g., the novelty of the experimenter). Accordingly, in subsequent testing sessions at all three field sites, the response period was extended to give young children additional time to process the prompt: after the initial pause, E2 repeated the anticipatory prompt a second time and then paused a second time. At all sites, the response period thus included the first pause, the second prompt, and the second pause. The exact lengths of these pauses varied across sites.

Below, we list additional modifications at the different sites.

Salar. In addition to the shared site variations, the procedure used with the Salar participants differed from that in Illinois in three ways. First, E1 did not offer the child a choice of paper or write the child's name on the paper. Second, before leaving the room, E1 also put away the child's paper and sticker, to minimize distractions. Third, during the first anticipatory prompt and response period, E2 looked down and covered her mouth with her hand (instead of looking at the ceiling, chin in hand); this was a culturally more appropriate way of delivering each prompt as a self-addressed question.

Shuar/Colono. Apart from the shared site variations, the procedure used in the Shuar/Colono communities differed from that in Illinois in that, before leaving the room, E1 put away the child's paper and sticker, to minimize distractions.

Yasawan. In addition to the shared site variations, the procedure used with the Yasawan children differed from that in Illinois in two ways. First, E2 put the scissors in her purse on the floor

(instead of in her pocket). Second, when delivering the prompt, E2 looked down at the floor to her right (instead of staring up at the ceiling, chin in hand); this was a more culturally appropriate way of delivering the prompt as a self-addressed question.

Figure S6. Verbal Anticipatory-Looking False-Belief Task being run at Shuar / Colono field site.

4.4 Coding

As mentioned earlier, the response period at all sites included the first pause after the initial anticipatory prompt, the second anticipatory prompt, and the second pause. Because pauses varied in duration across sites, when coding children's looking behavior the maximum length of each pause was set at 8 s (i.e., only the first 8 s of each pause were coded). The average duration of the response period was similar for the Salar (12.46 s) and Yasawans (14.23 s), but was longer for Shuar/Colonos (22.94 s), mainly because each pause lasted at least 8 s.

For each response period, we coded frame-by-frame whether the child was looking at the target container, the non-target container, or away. Our analyses focused on looking times at the target and non-target containers, rather than on looking times at the target container and at E2 (the scissors' initial and current locations), because children in this task typically look longer at E2 than at the target container during the response period: E2 is much more interesting and thus attracts more attention. This was true in the Illinois sample as well, making it necessary for He at al. *(21)* to compare looking times across the false- and true-belief conditions separately for each location. In these analyses, children in the false-belief condition looked longer at the target container than did those in the true-belief condition, whereas children in the true-belief condition looked longer at E2 than did those in the false-belief condition. He et al. also found that, whereas

children in the true-belief condition tended to look equally at the target and non-target containers, children in the false-belief condition looked reliably longer at the target than at the non-target container. This last finding guided our analyses here, since children at the field sites were tested only in the false-belief condition.

All response periods were coded independently by a second, naive coder. To ensure that this second coder was blind to which container was the target container, videos were edited to start at the onset of the test phase. Thus, the second coder had no information about whether the scissors were initially hidden in the left or the right container. Reliability was calculated as the percentage of frames in which the primary and secondary coders agreed on the child's direction of gaze (target container, non-target container, or away). Reliability was high across all three sites: Salar, 92%; Shuar/Colono, 95%; and Yasawan, 89%. Averaged across sites, the two coders agreed on the children's direction of gaze for 93% of the coded video frames.

4.5 Results

Figure 2 (see main paper) shows the mean looking times of children in each field site at the target and non-target containers during the response period. Looking time to the target container was longer than looking time to the non-target container for the Salar (target $M = 1.38$, $SD = 0.96$; non-target $M = 0.42$, $SD = 0.90$), the Shuar/Colonos (target $M = 1.03$, $SD = 1.15$; non-target $M = 0.32$, $SD = 0.66$), and the Yasawans (target $M = 1.37$, $SD = 1.32$; non-target $M = 0.75$, $SD = 0.97$). Because some of the distributions violated normality assumptions (a standard deviation much larger than the mean indicates a long right tail), we analyzed the data in two ways. First, we used one-sample t-tests against chance (0) to evaluate target advantage (looking time at target minus non-target container) during the response period; children in all three populations looked reliably longer at the target container: Salar, $t(12) = 2.94$, $p = .012$; Shuar/Colono, $t(28) = 3.27$, $p = .003$; and Yasawan, $t(6) = 2.85$, $p = .029$. Second, non-parametric Wilcoxon signed-rank tests also revealed longer looking at the target than at the non-target container: Salar, $Z = -2.41$, $p = .016$; Shuar/Colono, $Z = -3.00$, $p = .003$; and Yasawan, $Z = -2.02$, $p = .043$. Together, these results suggest that children represented E1's false belief and anticipated that she would look for her scissors in the target container.

An omnibus ANOVA with population, sex, and side of container as independent variables, and looking time to target and non-target containers as repeated measures, confirmed that looking times to the target container were significantly longer than those to the non-target container across populations ($F(1,37) = 12.9$, $p = .001$), with no main effects of the other variables (there were small interactions between sex and side ($F(1,37) = 4.63$, $p = .038$), and between population, sex, and side ($F(2,37) = 4.89$, $p = .013$)). An omnibus regression on all the populations pooled, with looking time difference (target minus non-target) as the dependent variable and population (dummy coded), sex, age (in months), and side of container where scissors were initially located (left or right) as independent variables, found no effects of any of these variables on looking time. Across the three populations, 71% (35/49) of the children looked at the target container during the response period, and 67% (33/49) looked longer at the target than at the non-target container; both percentages are comparable to those obtained by He et al. *(21)* in Illinois, which were 74% (14/19) and 68% (13/19), respectively.

Further Results: Older Shuar children. As noted earlier, pilot data were collected with a sample of Shuar children using a shorter response period that included only the 8-s pause following the first anticipatory prompt. With this shorter response period, positive results were obtained with an older age group of 5- and 6-year-olds ($n = 14$, 6 male and 8 female, age range: 59.7-73.0 months, $M = 64.6$ months). Within this age range, two additional children were tested but excluded because they were distracted by noises outside the testing room. During the response period, looking time to the target container ($M = 0.56$, $SD = 0.84$) was longer than looking time to the non-target container ($M = 0.04$ s, $SD = 0.13$). As before, we analyzed the data in two ways. First, a one-sample t-tests against chance (0) was used to evaluate target advantage (looking time at target minus non-target container) and showed that children looked reliably longer at the target than at the non-target container, $t(13) = 2.30$, $p = .039$. Second, a Wilcoxon signed-rank test also revealed longer looking at the target than at the non-target container, $Z = -2.67$, $p = .008$.

5. Non-Verbal Violation-of-Expectation False-Belief Task

5.1 Participants

In the Illinois sample *(25)*, participants in the non-verbal violation-of-expectation false-belief task were 18- to 19-month-olds ($M = 18.7$ months). To maximize the number of children that could be tested at each site, a broader age range was used that included children between 15.5 and 31 months.

Salar participants were 19 infants, 11 male and 8 female, ranging in age from 15.5 to 30.2 months ($M = 22.6$ months). Another 3 infants were tested but excluded, 2 because they were fussy and 1 because he was distracted by a loud noise in the test trial. Eleven infants (6 male) saw the different-object event, and 8 infants (5 male) saw the identical-object event.

Shuar/Colono participants were 19 infants, 14 male and 5 female, ranging in age from 16.9 to 30 months (M = 22.4 months). Another 6 infants were tested but excluded, 3 because they were fussy, 2 because they were inattentive, and 1 because his test looking time was over 3 standard deviations from the mean of his condition. Eleven infants (9 male) saw the different-object event, and 8 infants (5 male) saw the identical-object event.

Finally, 14 Fijian infants were tested, but the procedure did not produce usable data (see below).

5.2 Procedure in Illinois

Infants sat on a parent's lap in front of a display booth. Two naïve observers hidden on either side of the apparatus monitored each infant's looking behavior. The looking times registered by the primary observer were used to determine the ending of trials.

Infants received a familiarization trial and a test trial. At the start of the familiarization trial, an experimenter (E1) knelt at an opening in the right wall of the booth. Three objects (cups covered with lids) stood on the apparatus floor. One object stood directly in front of E1 (E1's object); the object was red and decorated with stars. Two additional objects stood at the rear of the apparatus: one matched the E1's object in pattern and color (identical object) and the other was green with

stripes (different object). To start, E1 grasped her object and shook it three times, demonstrating that it made a rattling sound. E1 then shook the different object, which also rattled, and then she shook the identical object, which made no noise. E1 repeated these actions, shaking each object in turn, until the child either looked away or looked a maximum of 60s. At this point, the familiarization trial ended (a curtain was lowered in front of the apparatus between trials).

Next, infants received a test trial in which E1 again knelt at the right window; a second experimenter, E2, was now present at the rear of the apparatus, behind the identical and different objects. To start, E1 grasped her object and lifted it. She then turned to E2 and said, "Look!" and shook her object as before. E1 then asked E2, "Can you do it?". E1 then put her object down and looked down at it. E2 then grasped either the identical (identical-object event) or the different (different-object event) object and paused. Children watched this paused scene until they either looked away or looked a maximum of 30 s. Half of the children saw the identical-object event and half saw the different-object event. For additional details of the procedure and descriptions of control conditions, see *(25)*.

5.3 Adaptation to field sites

At each site, infants were seated at the long edge of a table on the lap of a caretaker who was known to them. E1 was seated at one of the shorter edges of the table. E1's object stood in front of E1 and the other two objects stood across from the infant. The objects were closed, culturally appropriate containers; E1's object and the different object contained smaller objects (e.g., dried beans, pebbles) that produced a rattling sound when the object was shaken.

Because infants were tested in a house rather in a laboratory, two major modifications had to be made to the procedure. First, it was impossible for an observer to monitor the infants' looking behavior in real time during the testing session without being seen by the infants and drawing their attention. Instead, infants were videotaped and their responses were coded offline (see Coding section, §5.4). However, this change meant that the experimenters could not know whether the infants had reached the maximum looking time during the familiarization and test trials. This difficulty was overcome differently for the two portions of the experiment. The familiarization trial ended when either (1) the E1 had shaken each of the objects four times or (2) an observer seated behind the infant determined that the infant had become disinterested in the display. In the test trial, E1 and E2 remained paused until the observer behind the infant signaled that the infant had become disinterested in the display. In most cases, this was well after the infant had met the end of trial criteria (except for the Fijian children, see below). Second, the transition between the familiarization and test trials had to be modified. At each site, E2 waited outside the room during the familiarization trial. Once this trial was completed, E2 was signaled to enter the room. E1 paused and looked to a neutral location until E2 was in position, then continued on with the test trial.

Below, we list additional modifications at the different sites.

Salar. In addition to the shared site variations, the procedure used for the Salar differed from that in Illinois in two ways. First, the colors of the objects were counterbalanced. For half of the infants, E1's object and the identical object were red while the different object was green. For the

other half of the infants, these colors were reversed. In both cases, E1's object and the different object made noise when shaken, but the identical object did not. Second, infants received two familiarization trials instead of one. This was done because the research assistants on site were concerned that the procedure would be very novel to the infants and thus they would require more time to learn the objects' properties. The observer behind the infant used a computerized sound to signal the start and end of each trial, and coughed to tell E2 to enter the room.

Shuar/Colono. In addition to the shared site variations, the procedure used in the Shuar/Colono communities differed from that used in Illinois in one way: whether the E1 sat to the infant's left or right varied across infants. Regardless of where the E1 sat, his object was placed directly in front of him and the different/identical objects were placed across from the infant with the identical object furthest from E1. In the familiarization phase, E1 always shook his object, then the different object, and then the identical object. The observer behind the infant coughed to tell E2 to enter the room, and stood up to signal when the test trial should end.

Fijian. Rather than waiting until the infant had become disinterested in the test display, the observer in the Fijian community attempted to guess when the infant had met the end of trial criteria (see Coding section, §5.4). However, the observer misunderstood what counted as looking at the display, incorrectly ending the trial if the infant looked at the experimenters. As a result, for 11 of the 14 infants tested, the test trial ended while the infants were still actively attending to the display. Thus, there was no usable data for this experiment at this field site.

Figure S7. Non-Verbal Violation-of-Expectation False-Belief Task being run at Salar field site.

5.4 Coding

Infants' looking behavior was coded from video. Infants were coded as looking if they were attending to an area that included both the experimenters and the objects. All other areas were coded as away.

Familiarization trial. For the Salar infants, we coded from the onset sound until the computer sounded that the trial had ended. For the Shuar/Colono infants, there was no onset and offset sound. Instead, we coded from the time E1 touched the first object until the time the E1 released the last container on the fourth rotation. Because there were two familiarization trials for the Salar infants, but only one for the Shuar/Colono infants, we summed the Salar infants' looking times for the two trials into a single trial.

Test trial. In the test trial, we coded the pretrial and the paused scene separately for both populations. For the Salar, the pretrial began at the trial onset sound, while for the Shuar/Colonos it began when E1 touched E1's object. At both sites, the pretrial included E1's demonstration as well as E1's prompt. The pretrial ended when E2 touched a container, at which point the paused scene began. For the paused scene, we coded until the infant had either (1) looked away for 2 consecutive seconds after having looked for at least 8 cumulative seconds or (2) looked for 30 cumulative seconds.

All test trials were coded independently by a second, naïve coder. To ensure that this second coder was blind to the participant's condition, the videos were edited to start at the onset of the paused scene and were cropped so that the coder could not see which container E2 reached for. Reliability was calculated as the percentage of frames in which the primary and secondary coders agreed on the child's direction of gaze (at the display or away). Reliability was high for both the Salar (95%) and Shuar/Colono (93%) participants. Averaged across sites, the two coders agreed on the infants' direction of gaze for 94% of the coded video frames.

Preliminary analyses revealed no effects of container color (Salar only). The data were therefore collapsed across this factor in subsequent analyses.

5.5 Results

Familiarization trial. In each population, we compared the looking times during the familiarization trial of the infants who saw the identical- or the different-object event in the test trial. Infants' responses were analyzed using independent-samples t-tests, which revealed no significant differences, both ts < 1. Next, we conducted an omnibus ANOVA on both populations pooled with test event (identical- or different-object), sex, and population as between-subjects factors (Note: The small number of females in the Shuar/Colono group prevented us from conducting separate ANOVAs for each country). The analysis revealed no significant effects, all Fs < 1.19, all ps > .28. Finally, an omnibus linear regression found no effects of test event, sex, population or age on looking time. Thus, all groups of infants were equally attentive during familiarization.

Test trials. Figure 3 (see main paper) shows the infants' mean looking times during the test trial,

separately by test event and field site. Infants' responses were analyzed using independent-samples t-tests. In each population, infants who were shown the different-object event looked reliably longer than those shown the identical-object event: Salar $t(17) = 2.25$, $p = .038$ (different $M = 17.94$, $SD = 7.38$; identical $M = 11.57$, $SD = 3.49$); Shuar/Colono $t(17) = 2.13$, $p = .048$ (different $M = 19.84$, $SD = 7.08$; identical $M = 14.09$, $SD = 3.26$).

An omnibus ANOVA with event (identical- or different-object), sex, and population as between-subjects factors confirmed the main effect of event, $F(1, 30) = 7.72$, $p = .009$. This analysis also revealed a main effect of sex, $F(1, 30) = 7.02$, $p = .013$, which resulted from the fact that females ($M = 12.50$) tended to look relatively less than males ($M = 18.33$). Finally, the analysis also revealed an interaction of sex and event, $F(1, 30) = 4.40$, $p = .045$, resulting from the fact that the difference in looking time at the different- and identical-object events was larger for the males (different $M = 21.75$; identical $M = 13.20$) than for the females (different $M = 12.75$; identical $M = 12.20$). Since there were many fewer females (13) than males (25) in the total sample, these differences should be interpreted with caution.

Finally, an omnibus linear regression found a significant effect of event ($\beta = .42$, $p = .004$) and a significant effect of sex ($\beta = .40$, $p = .005$) on looking time. There was no significant effect of population or age (in months).

6. Additional field site: Kenya

The verbal preferential-looking and anticipatory-looking tasks were also administered to children at a field site in the eastern region of Kenya near the Rift Valley in the Bungoma district. Participants were recruited by word of mouth from rural villages near Bungoma town, and tested in the center of Bungoma town. Unfortunately, the data collected at this field site were unusable and were therefore excluded from the analyses.

Verbal preferential-looking task. This task depends on children's tendency to look longer at images that match the sentences they hear. Like the Illinois children, the Salar, Shuar/Colono, and Yasawan children demonstrated this basic tendency in the setup trials by looking longer at the matching than at the non-matching picture. However, the children in Kenya looked about equally at the matching and non-matching pictures during the setup trials, suggesting that they were not following the story they were told. Because children gave no evidence that they were following the story during the setup trials, their behavior during the test trial was uninterpretable.

Verbal anticipatory-looking task. This task depends on differential looking-times at the target container (where E1 finds and replaces her scissors) and the non-target container. Unfortunately, we had a small subject population for this task in Kenya (18), and the majority (11) of the children failed to look at either container during the response period.

Why did the children at the Kenya field site fail at our tasks? Although we cannot know for certain, one possible explanation is that children in this region tended to attend to the social demands of the testing situation rather than to the task itself. In another empirical study conducted in the same region, children 1-6 years of age responded with inhibition on a mirror self-recognition task: all but two of the children failed the task, even though these same children

were seen inspecting themselves in mirrors during natural observations *(57)*. The interpretation of these results, and of those under discussion here, is that the children were inhibited by the testing situation and the social expectations about their performance, and as a result they gave relatively less attention to the task itself.

7. References

38. Wellman, H.M., Fang, F., Liu, D., Zhu, L., Liu, G. 2006 Scaling of theory-of-mind understandings in Chinese children. *Psych. Sci.* 17, 1075-1081.

39. De Villiers, J.G., Pyers, J.E. 2002 Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cog. Dev.* 17, 1037-1060.

40. Dunn, J., Brown, J., Slomkowski, C., Tesla, C., Youngblade, L. 1991 Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Dev.* 62, 1352-1366.

41. Ruffman, T., Slade, L., Crowe, E. 2002 The relationship between children's and mothers' mental state language and theory-of-mind understanding. *Child Dev.* 73, 734-751.

42. Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. 2005 Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–735.

43. Harris, P.L. 2005 in *Why Language Matters for Theory of Mind*, eds Astington JW, Baird JA (Oxford Univ. Press, New York), pp 70-83.

44. Robbins, J. 2008 On not knowing other minds: Confession, intention, and linguistic exchange in a Papua New Guinea community. *Anth. Quart.* 81, 421-429.

45. LeVine, R.A. 1984 in *Culture Theory: Essays on Mind, Self, and Emotion,* eds Shweder RA, LeVine RA, Eds. (Cambridge Univ. Press, New York), pp 67-87.

46. Danziger, E. 2006 in *Roots of Human Sociality: Culture, Cognition, and Interaction,* eds Enfield NJ, Levinson SC (Berg, New York), pp 259-278.

47. Duranti, A. 1988 Intentions, language, and social interaction in a Samoan context. *J. Prag.* 12, 13-33.

48. Duranti, A. 2008 Further reflections on reading other minds. *Anth. Quart.* 81, 483-494.

49. Lancy, D.F. 1996 *Playing on the Mother Ground: Cultural Routines for Children's Development* (Guilford, New York).

50. Astington, J.W., Baird, J.A. 2005 Eds., *Why Language Matters for Theory of Mind.* (Oxford Univ. Press, New York).

51. Harner, M. 1972 *The Jivaro: People of the Sacred Waterfalls.* (Univ. California Press, Berkeley, CA).

52. Hagen, E.H., Barrett, H.C. 2009 Cooperative breeding and adolescent siblings: Evidence for the ecological constraints model? *Curr. Anth.* 50, 727-737.

53. Everett, D.L. 2005 Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Curr. Anth.* 46, 621-646.

54. Broesch, T., Olah, K., Rochat, P., Mother-infant interactions in three cultures. (in preparation).

55. LeVine, R.A., Dixon, S., LeVine, S., Richman, A., Leiderman, P.H., Keefer, C.H., Brazelton, T.B. 1994 *Childcare and Culture: Lessons from Africa*. (Cambridge Univ. Press, New York).

56. Morton, H. 1996 *Becoming Tongan: An Ethnography of Childhood*. (Univ. Hawai'i Press, Honolulu, HI).

57. Broesch, T., Callaghan, T., Henrich, J., Murphy, C., Rochat, P. 2011 Cultural variations in children's mirror self-recognition. *J. Cross-Cult. Psych.* 40: 1019-1031.