Retrospective attribution of false beliefs in 3-year-old children

Ildikó Király[a,b,1], Katalin Oláh[a], Gergely Csibra[b,c], and Ágnes Melinda Kovács[b]

[a]MTA-Momentum Social Minds Research Group, Eötvös Loránd University, 1064 Budapest, Hungary; [b]Department of Cognitive Science, Central European University, 1051 Budapest, Hungary; and [c]Birkbeck, University of London, Bloomsbury, WC1E 7HX London, United Kingdom

Edited by Renée Baillargeon, University of Illinois at Urbana–Champaign, Champaign, IL, and approved September 17, 2018 (received for review February 27, 2018)

A current debate in psychology and cognitive science concerns the nature of young children's ability to attribute and track others' beliefs. Beliefs can be attributed in at least two different ways: prospectively, during the observation of belief-inducing situations, and in a retrospective manner, based on episodic retrieval of the details of the events that brought about the beliefs. We developed a task in which only retrospective attribution, but not prospective belief tracking, would allow children to correctly infer that someone had a false belief. Eighteen- and 36-month-old children observed a displacement event, which was witnessed by a person wearing sunglasses (Experiment 1). Having later discovered that the sunglasses were opaque, 36-month-olds correctly inferred that the person must have formed a false belief about the location of the objects and used this inference in resolving her referential expressions. They successfully performed retrospective revision in the opposite direction as well, correcting a mistakenly attributed false belief when this was necessary (Experiment 3). Thus, children can compute beliefs retrospectively, based on episodic memories, well before they pass explicit false-belief tasks. Eighteen-month-olds failed in such a task, suggesting that they cannot retrospectively attribute beliefs or revise their initial belief attributions. However, an additional experiment provided evidence for prospective tracking of false beliefs in 18-month-olds (Experiment 2). Beyond identifying two different modes for tracking and updating others' mental states early in development, these results also provide clear evidence of episodic memory retrieval in young children.

theory of mind | episodic memory | memory development | prospective mind reading | retrospective mind reading

Humans are undoubtedly ultrasocial beings: they live their lives in an almost continuous flow of interactions (1). This ubiquitous sociality imposes an enormous sociocognitive demand: To engage in communication, collaborations, or any event that is governed by socially formed concepts, such as norms or customs, they need to be able to take into account the mental states of their social partners. Accordingly, everyday functioning requires humans to become experts in monitoring others' minds to predict and interpret their behavior and their utterances, an ability also termed theory of mind (ToM).

Although an abundance of studies has investigated whether, and under what circumstances, children and adults attribute mental states, there is only scarce empirical evidence regarding the processes involved in how people compute and dynamically update attributed mental states in real time. In the typical paradigms used for testing ToM competencies, the participant is exposed to the following event sequence: a character, Sally, puts her marble into a basket and leaves. Another character, Anne, moves the marble to a box before Sally returns for her marble. In the explicit version of the task, at this moment, participants are prompted to answer direct questions regarding Sally's impending action, which require them to take into account her beliefs about the situation (2, 3). Young children usually fail to answer these questions correctly, but implicit versions of this task, developed in the last 2 decades, have provided ample evidence that infants, similarly to adults, can track a character's beliefs, even without being asked to do so, as reflected by their looking times (4–6), anticipatory looks (7, 8), or active behavior (9, 10). However, whether the tasks were implicit or explicit, previous studies relied on protocols that did not allow disentangling the different cognitive mechanisms required for passing the tasks.

In particular, it is unclear how and at what point of the event stream beliefs are computed and attributed in these tasks. Taking the above location-change paradigm as an example, belief attribution could take place at the beginning of the story, when Sally puts her marble into the basket; at the end, when the participant is prompted to predict Sally's behavior; or in between, for example, when Anne relocates the marble. Crucially, these options differ not only in timing but also in inferential and computational requirements: Different functional mechanisms may be recruited at the different points of the scenario. Attributing a (true) belief at the beginning of the story, when the protagonist's perceptual access to a state of affairs is encoded, requires the maintenance of this attributed belief, despite changes of reality, as the events unfold, to succeed in the task (11). This can be termed prospective belief attribution because it may or may not have any immediate use for the observer, but the attributed belief can be stored and maintained in case it is required in further inferences. Such a prospective mechanism of belief attribution does not even have to track the truth-value of the belief for enabling passing a false-belief (FB) task, and does not necessarily require encoding the source event that led to belief attribution.

In contrast, if belief attribution takes place at the end of the story, when the content of the relevant belief is needed for action

Significance

The continuous flow of social interactions requires humans to monitor others' mental states dynamically, yet this aspect of mind reading remains largely neglected. We tested whether, beyond prospective belief tracking, young children would also attribute beliefs to others retrospectively. We found that 3-year-old children retrospectively inferred the content of someone's beliefs by combining present information with relevant events retrieved from episodic memory. This finding shows that emerging capacities for episodic memory contribute to the development of social cognitive processes, enriching children's ability to monitor others' mental states.

Author contributions: I.K., G.C., and Á.M.K. designed research; I.K., K.O., and Á.M.K. performed research; I.K., K.O., G.C., and Á.M.K. analyzed data; and I.K., K.O., G.C., and Á.M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: kiralyi@caesar.elte.hu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1803505115/-/DCSupplemental.

prediction, computing the content of the belief must be based on a memory search targeting all relevant information that can potentially contribute to the identification of such content. This search may be triggered spontaneously in implicit tasks (e.g., by the reappearance of the actor whose belief is relevant; i.e., when Sally returns to find her marble), or by the direct question regarding the protagonist's beliefs or actions in explicit tasks. Although this retrospective mechanism of belief attribution does not require continuous tracking and maintenance of attributed beliefs, it can only be performed successfully if all relevant details of past events are faithfully preserved and accessible when needed. For instance, to pass (explicit or implicit version of) the Sally–Anne task by retrospective belief attribution, one should recall the episode when Sally put her marble into the basket, trace the intervening events related to her and/or the marble (e.g., that she did not see the transfer), and infer that Sally still believes her marble is in the basket.

In essence, these routes of belief attribution also presume differences in how changes to attributed beliefs are traced. In prospective tracking, attributed beliefs can be updated: an initially ascribed and maintained belief can be modified when a new belief-relevant event occurs that necessitates such a change. There is emergent evidence that children can perform such updates with respect to their own representations (12), and perhaps also with respect to attributed beliefs (13, 14). For example, infants in the second year of life understand that an actor's FB about the location of an object can be corrected by communication (13, 14). Thus, in prospective belief tracking, observers monitor relevant events continuously, and as soon as the belief of the agent should change, they also perform the appropriate updates.

However, retrospective, memory-based belief inferences would also allow people to retrieve and revise information that may not have been taken into account in prospective belief attribution. Such retrospective revisions are initiated when some information indicates that a belief might have been incorrectly attributed and must be revised. In contrast to the prospective updates, such revisions correct one's own earlier inference (what one thinks about an agent's belief, whereas the agent's belief itself is not thought to change), and require the retrieval of episodic memories of the sources of this inference. Thus, episodic retrieval is an important mechanism for retrospective belief revision: in case novel information comes up regarding the past context that induced prospective belief attribution, one can retrospectively recompute the content of already attributed beliefs (15).

These two mechanisms of belief attribution and belief tracking are not mutually exclusive, but may work in an integrated manner. If, for example, Sally's belief is attributed when Anne relocates the marble, it might be based on retrospective recalling of what happened before (Sally saw the marble in the basket), and the resulting belief should be prospectively maintained until it is exploited for action prediction. Importantly, in the common FB tasks, these computational strategies cannot be disentangled because they predict similar outcomes. It is possible that successful performance in these tasks is simply based on prospective belief attribution and maintenance of these attributed beliefs throughout the event. In fact, because retrieving past episodes poses difficulty for young children (16, 17), it is a plausible assumption that their successful performance in implicit FB tasks relies on prospective, rather than retrospective, attributing mechanisms. The purpose of the present study was to test whether and when retrospective attribution mechanisms are available to children in implicit tasks. To achieve this aim, we had to develop a task that cannot be solved by purely prospective belief computations.

Experiment 1

We developed a paradigm relying on the referential disambiguation ToM task of Southgate et al. (18). The crucial manipulation we introduced was a belief revision phase in between the belief induction phase and the test. The task had the following structure: In the belief induction phase, an experimenter (E1) hid two novel objects into two boxes; later, while wearing sunglasses, she "witnessed" as the location of the two objects were swapped. This scene may lead to the prospective attribution of a true belief (TB) to E1 about the respective location of the objects, given that sunglasses are usually transparent. In the belief revision phase, while E1 was away, the participants explored her sunglasses, which turned out to be either opaque or transparent. In the condition where the sunglasses were opaque, to correctly track E1's belief, children had to retrospectively revise its status from TB to FB and recompute its content regarding the location of the objects. We label this condition TB-FB, indicating that, to succeed, children had to retrospectively change the status of the attributed belief from true to false. In the condition where the sunglasses were transparent, retrospective revision of the attributed belief was not necessary (TB-TB condition).

In the subsequent test phase, E1 returned, pointed to one of the boxes, and asked for an object. The dependent measure of the study was whether children, in response to this request, gave her the object from the referred or from the other (nonreferred) box. In line with the original study (18), we built our predictions on the following consideration: when the experimenter pointed to the box containing an object, children would not interpret the gesture as referring to the box itself; rather, they should map it to the object hidden inside. Importantly, this referent mapping is dependent on the attributed belief: E1's gesture must refer to the object she (truly or falsely) believes to be located in that particular box. If children in our task retrospectively revise the TB to a FB when the sunglasses turn out to be opaque during the belief revision phase (TB-FB condition), they should choose the nonreferred box. In contrast, children are expected to choose the referred box in the TB-TB condition, when such a revision was not required.

We tested whether these memory-based retrospective belief attribution mechanisms were available for 18- and 36-mo-old children. The younger group represents an age when infants have been shown to pass interactive FB tasks (9, 10, 18, 19), and the older group targeted the age when episodic memory capacities seem to emerge (20, 21).

The number of infants who chose the referred and the nonreferred box in each condition is depicted in Fig. 1. The infants in the younger age group preferred to give the experimenter the referred object in both conditions: Twelve infants chose the referred box and 4 infants chose the nonreferred one in the TB-TB condition, and 11 infants chose the referred box and 4 infants chose the nonreferred one in the TB-FB condition. In contrast, the 36-mo-olds displayed a different pattern across the two conditions: in the TB-TB condition, 16 participants chose the referred box and 2 chose the nonreferred one, whereas in the

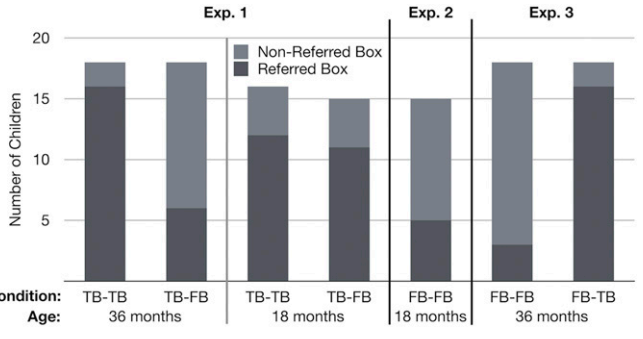

Fig. 1. Number of children choosing the referred or nonreferred box as a function of condition and age in Experiments 1, 2, and 3.

TB-FB condition, only 6 children chose the referred box and 12 chose the nonreferred one.

A 2 × 2 (age × condition) log-linear analysis revealed that the pattern of answers across the conditions differed in the two age groups significantly ($G^2 = 13.98$; $df = 4$; $P < 0.01$). Fisher's exact tests show that although in 18-mo-olds there was no difference between the conditions ($P = 1.000$), in the 36-mo-old sample, the number of children choosing the referred box differed significantly between the TB-TB and TB-FB conditions ($P = 0.002$). This analysis targeted our main question; specifically, whether the response pattern in the TB-FB condition in the two age groups shifted away from that of the TB-TB condition. Children's natural (baseline) response to a request accompanied by a pointing gesture is not a random choice between the available options, but it is better represented by the pattern they produced in the TB-TB condition.

These results revealed no evidence that the 18-mo-olds would have considered their experience with the sunglasses as relevant to their response to E1's request. However, the 36-mo-olds behaved differently in the two conditions, suggesting that they were able to identify that, in the TB-FB condition, the information revealed about the opacity of the sunglasses during the belief revision phase was relevant for E1's belief state. As a consequence, they must have recalled that E1 had been wearing sunglasses during the location change event, retrospectively recomputed her belief about the location of the objects, and used this information to respond to her request.

The failure of the 18-mo-olds in the TB-FB condition might be a result of prospectively maintaining the already attributed belief during the belief revision phase (i.e., attributing TBs in both conditions). However, it is also possible that the infants simply followed the referential request of the model (by giving her the referred object in both conditions) without paying attention to her potential belief content in either condition. Experiment 2 investigated this question.

Experiment 2

Considering that Southgate et al. (18) found that 17-mo-olds resolved an ambiguous referential request by appealing to the (true or false) belief of their interlocutor, the 18-mo-olds' failure in Experiment 1 seems unlikely to be because of their inability to take into account FBs per se. However, one might argue that they did not consider the opacity of the sunglasses as causally relevant for the belief states of the actor (but see ref. 22). In Experiment 2, we tested whether information about the opacity of the sunglasses revealed before the encoding of E1's belief would lead infants to prospectively infer that she would have a FB. If 18-mo-olds pass this test, it would indicate that immaturity of belief revision mechanisms (rather than deficient belief attribution mechanisms) made them ignore this causally relevant information about the sunglasses in Experiment 1.

The infants were shown a pair of opaque sunglasses to explore during familiarization, and then E1 wore the very same sunglasses (identical to the ones introduced in Experiment 1) during the belief induction phase. Thus, in this experiment, the infants knew in advance that the sunglasses worn by E1 while observing the swap of the objects were opaque (hence we call it the FB-FB condition).

In Experiment 2, five 18-mo-old infants chose the referred box and 10 infants chose the nonreferred one (Fig. 1). The results of this experiment were compared with those of the two conditions with 18-mo-olds from Experiment 1. A 2 × 3 χ^2 test revealed a difference in the pattern of choices between the three conditions ($\chi^2 = 7.29$; $df = 2$; $P = 0.029$). Follow-up Fisher's exact tests confirmed that the number of infants choosing the referred box differed significantly between the TB-TB and FB-FB conditions ($P = 0.022$), and also between the TB-FB and FB-FB conditions (Fisher's exact $P = 0.033$). This pattern of results suggests that, in contrast to Experiment 1, infants could take into account the (false) belief state of E1 when responding her request. This finding essentially replicates that of Southgate et al. (18).

Experiment 3

In Experiment 3, we intended to provide a conceptual replication of the finding that 36-mo-olds revise an attributed belief retrospectively (Experiment 1). To this end, we reversed the direction of belief revision. If 36-mo-olds can perform flexible retrospective belief revision, this should be independent of whether the starting point is a TB or a FB. Thus, we tested whether 36-mo-olds revised an attributed FB when they subsequently learned that the agent could have witnessed the situation that they initially thought had not been perceived by her. We hypothesized 36-mo-olds would be able to infer that the agent should have a TB, rather than a FB.

During the experiment, children witnessed events that were similar to Experiments 1 and 2, except that E1 left the room while the location change of the objects took place. Children were then asked to accompany E2 to the other room to invite E1 back. When they entered the other room, children in the FB-TB condition observed E1 peeking into the experimental room through a one-way mirror. In contrast, when they collected E1 in the FB-FB condition, they did not receive such information (the mirror was covered). When E1 was back, she requested an object from the children the same way as in Experiments 1 and 2.

In response to this request, 15 participants chose the nonreferred box and three chose the referred one in the FB-FB condition. However, in the FB-TB condition, in which they witnessed E1 peeking through the one-way mirror, 16 children chose the referred box, and only two of them chose the nonreferred one (Fig. 1). Fisher's exact test confirmed that the number of children choosing the referred box differed significantly between the FB-TB and FB-FB conditions ($P = 0.0001$).

This pattern of results corroborates data from Experiment 1. It confirms that in the FB-FB condition, the 36-mo-olds attributed a FB to E1 and responded to her request accordingly. More important, the children in the FB-TB condition recomputed E1's belief about the location of the objects and attributed her a TB. This suggests that they detected the relevance of the one-way mirror and understood that E1 could have had visual access to the location change of the objects.

Discussion

Tracking what others know and believe plays an important role in human communication because utterances have to be designed in production, and interpreted in comprehension, in the context of the mental states of one's interlocutor (23). It has been previously demonstrated that infants can disambiguate referential expressions by appealing to the belief content of the communicator, even when this belief is false (18). We confirmed this finding in the FB-FB conditions of Experiments 2 and 3, in which children had information about the fact that E1 did not have perceptual access to the location change of the objects, and were able to use this in interpreting her referential expressions. Experiment 2 provided further evidence that 18-mo-olds are able to apply their self-experience (regarding the opacity of the sunglasses) to assess the experimenter's lack of visual access (22).

Importantly, however, in Experiments 1 and 3, 36-mo-olds (but not 18-mo-olds) displayed sophisticated belief revision capacities as well. In Experiment 1, they revised a TB into a false one without directly observing the communicator's lack of perceptual access generating her FB. In Experiment 3, they also succeeded in such a revision in reversed order, revising a FB into a TB, without directly witnessing that the communicator had perceptual access to the crucial events. Although the presence of an actor usually leads to TB attributions and the absence to false ones, the 36-mo-olds in our study could overcome this rule when needed. However, in contrast to earlier studies (13, 14), children

here did not prospectively update the actor's belief when its content changed, but corrected their earlier mistaken inferences regarding this belief. In other words, they performed a retrospective belief revision. Thus, at the age of 3 y, when young children do not normally succeed on explicit FB task (but see ref. 24), they demonstrate efficient retrospective belief revision abilities relying on episodic memories, recruiting complex belief computation processes that likely involve backtrack reasoning and reevaluating causal relations that lead to a particular belief, which can be difficult even for adults (25).

FB attribution in all three experiments required inferential processes, however: 36-mo-olds in Experiments 1 and 3 could not have succeeded in the task without invoking memories of specific past events into their inferences. In particular, at some point of the procedure, they must have recollected the location-swap event during the belief induction phase to combine this memory with the later acquired information about the sunglasses (Experiment 1) or one-way mirror (Experiment 3) to recompute E1's belief content. Thus, the results of our study cannot be explained by prospective attribution and updating mechanisms and demonstrate the operation of retrospective belief attribution mechanisms in 36-mo-olds.

Nevertheless, our findings leave open questions regarding the nature of these mechanisms. Here we consider three alternative scenarios, which differ in the assumptions as to when initial belief attribution took place and what events triggered these attribution processes. Belief attribution could have been initiated by the observation of E1's perceptual access regarding the location of the objects during the belief induction phase (or the lack of it), by the need to interpret her request in the test phase, or by acquiring information about the opacity of the sunglasses or the presence of the one-way mirror during the belief revision phase. These alternative scenarios assign different roles to the episodic memory-based retrospective processes that must have operated during the task to result in the correct responses that we observed in 36-mo-olds.

The first and most plausible way to think about retrospective attribution mechanisms is that they always operate on, and modulate, already-attributed (true or false) beliefs. According to this option, the primary mechanism of belief attribution is the prospective route: children always attribute (true) beliefs when they observe an agent's perceptual access to some relevant facts and then maintain these attribution across events, updating the belief content when it is necessary. Thus, children in the standard location-change FB tasks (and our participants in Experiments 2 and 3) initially attribute a TB about the object's location to the protagonist (E1 in our study), and then maintain this belief attribution when the content of the belief becomes false in the absence of the protagonist (as in Experiment 3), or in the lack of perceptual access of the experimenter (wearing opaque sunglasses in Experiment 2). In the present study, when 36-mo-olds learnt about the opacity of the sunglasses in the TB-FB condition of Experiment 1, and subsequently recomputed the content of the experimenter's belief, they did not attribute a new belief to her, but rather revised or corrected a prospectively attributed belief. To be able to do so, they must have stored not only the content of this belief but also some relevant facts about the source of this attribution (i.e., E1's perceptual access to the objects' location) and history of updates (e.g., when location change occurred within the stream of events). Then either at the point when they found out that the sunglasses were opaque or at the point when they had to evaluate E1's referential expression, they retrospectively revised the source information and the update history of the attributed belief in light of what they learnt about the sunglasses, and recalculated the content of E1's belief accordingly. Although they had to preform similar revisions (with opposing directionality) when they were exposed to the one-way mirror of the FB-TB condition of Experiment 3, there was no need for such a revision in the TB-TB condition of Experiment 1 or in the FB-FB of condition of Experiment 3.

If this is the correct interpretation of retrospective attribution, then the 18-mo-olds' failure in Experiment 1 was the result of the inaccessibility of source and update information either because of failing to retrieve it from memory or because of failing to encode it in the first place. In either case, these infants were unable to revise their prospective belief attribution and erroneously relied on it in the test phase.

Second, in principle it is possible that all responses in our tasks, and even in the majority of FB tasks, were based on purely retrospective mechanisms. If children possess sufficiently accurate mechanisms of recalling relevant details of past events, they could calculate the belief content of the communicator at the time when they need it (i.e., when they have to predict her action or interpret her request).

With respect to our findings, this option would assume that the children in all conditions calculated the content of E1's relevant belief when she pointed to a box and asked for an object. This belief attribution process would have been similar in the TB-FB/FB-TB and TB-TB/FB-FB conditions, except that the perceptual access of the experimenter to the location-change event had to be evaluated differently, depending on the subsequently acquired information concerning the transparency or opacity of the sunglasses and the presence of the one-way mirror. If this interpretation were correct, the 18-mo-olds' failure in Experiment 1 would not be a result of memory limitation or to the absence of retrospective attribution mechanisms, but an inferential deficiency: They would not have recognized that the information they learnt about the sunglasses or the one-way mirror was relevant to the evaluation of E1's perceptual access to the earlier event that fixed the belief content.

Although on the basis of our study we cannot exclude this explanation, we find this option, which completely eliminates prospective attribution processes, unlikely: building a belief attribution system entirely on retrospective mechanisms would require highly reliable, fast, and accurate retrieval of events from episodic memory. Considering that even adults' episodic memory fails to meet these standards (26, 27), and young children's ability to recall past events is much weaker than that of adults (28–30), the wide range of findings on early belief attribution calls for mechanisms that should not exclusively depend on memory.

A third possible way of interpreting the role of retrospective belief attribution in theory of mind is to link it to belief update mechanisms that operate on one's own beliefs. According to this option, an initial prospective TB attribution is not mandatory, and both prospective and retrospective attribution and revision mechanisms are triggered when the contents of one's own beliefs are updated or revised. In the standard, location-change FB task, the informational access of the protagonist to the location of the marble does not have to be recorded by creating a separate (meta)representation; it can be represented by simply tagging the child's own representation of the location by the protagonist, indicating that she has access to this information (31). When the marble is relocated and content of this representation is updated accordingly, the tagging is also updated. If the protagonist has perceptual access to the change, the tagging is maintained on the representation; if she does not have access, a new representation is created and attributed to her with the old content. This latter process is a prospective attribution of a FB, triggered not by perceptual access but by the lack of it. This form of belief attribution would thus have both retrospective and prospective elements, and although it is not a revision of an already-attributed belief, it is triggered by the revision of own beliefs, also taking into account the perceptual access of the protagonist.

Such prospective FB attribution would explain infants' success in the FB-FB task of Experiment 2 (as well as in ref. 18). However, in Experiment 1, children had to update their representation of current reality not only when the locations of the objects were swapped but also when they learned that the sunglasses, which they initially believed to be transparent, were opaque. This revision process

might have triggered the search for additional representations, linked to the updated information, to be revised. Indeed, it has been suggested that one function of episodic memory is to allow us to revise our own beliefs on the basis of new information related to the original source of those beliefs (15). Such a search might have led the 36-mo-olds in Experiment 1 to memories related to E1, who had previously worn the sunglasses, and might have allowed them to retrospectively reevaluate her perceptual access to the location change event. As a consequence, they could remove E1's tag from their own representation of true reality and could create a new representation by attributing to her the FB with the content of the location of the objects before the swapping took place. In Experiment 3, having learnt about the one-way mirror, children could revise the attributed belief the same way as described in the first option above, and such a revision would be triggered by learning about the presence of the one-way mirror.

If this is the right interpretation of the results, 18-mo-olds might have failed to attribute a FB in Experiment 1 because their search for to-be-revised information related to the opacity of the sunglasses did not lead them to the memory of the particular event during which the experimenter had worn those sunglasses. In other words, weak or unreliable memory traces, or immature recollection processes, could explain their failure.

The question of which of these alternative accounts explains our findings best is beyond the scope of this article and will have to await for further investigations. Nevertheless, our study demonstrates that retrospective belief attribution mechanisms are available to children from at least 3 years of age, which raises the question of whether all findings in the relevant literature on mind reading are to be explained by purely prospective mechanisms. In addition, this pattern of results suggests that updating and revision of attributed beliefs relies on flexible manipulation of representations and metarepresentations even in so-called implicit ToM tasks and at an age at which children have been claimed to lack proper ToM abilities (32). Such a conclusion is inconsistent with views that explain young children's performance in such tasks by associative learning (33) or by tracking agent-object relations (32), as these approaches do not include options for combining belief-relevant information originating from different sources, such as episodic memory.

Furthermore, our study provides evidence of episodic retrieval processes, which are necessary for retrospective belief revision, at the age of 36 mo. There is a growing body of evidence that 3-year-old children perform well on certain episodic memory tasks (21, 34–36). For example, in a typical task (21), children first dig up a locked treasure-case from a sandbox, and later in another room they are allowed to select one of three items (a key and two distractor objects) to take back to the sandbox scene. Three-year-olds chose the key if the delay between the events was 15 min or less. Although such results demonstrate that children can use information acquired during a past event for an upcoming event, this achievement can be based on carrying over some semantic information extracted from that past event ("sandboxes have locked treasure cases") or on tracking current states of affairs ("there is a locked treasure case in that sandbox") without retrieving the details of a specific past event.

In contrast to such tasks, our paradigm required children to recall episodic information related to E1 and a specific event, which could not have been achieved by recalling semantic information about her, about sunglasses or one-way mirrors, or about object location, or by tracking states of affairs concerning these elements separately. In other words, having learnt that the sunglasses were opaque or that E1 had visual access to the experimental room, children had to retrieve the specific event when E1 had worn those glasses or what had happened in that room, and only within the frame of the original event could they infer the consequences of seeing or not seeing the location change. This retrospective attribution process is in line with proposals that episodic memories enable updating the inferential consequences of past events in light of newly acquired information (15). The present study thus provides clear evidence that 36-mo-olds can recollect episodes (at least within a minute delay) in sufficient detail to be used for revising their inferences regarding the beliefs of interacting partners.

Methods

The experiments were approved by the ethical committee of Eötvös Loránd University. Parents of all participants signed an informed consent form before starting the experiments.

Experiment 1.

Participants. The planned sample size was 40 children for each age group (18- and 36-mo-olds), equally distributed to the two conditions. Three 18-mo-olds and a single 36-mo-old were excluded and replaced because of experimenter error. Some children did not make a choice or chose both boxes during the test: nine 18-mo-olds (TB-TB condition: four, TB-FB condition: five) and four 36-mo-olds (TB-TB condition: two, TB-FB condition: two). Because these participants completed the task, they were not replaced. Thus, the final sample that produced evaluable data included 31 18-mo-olds (TB-TB condition: 16; TB-FB condition: 15; mean age, 18.1 mo; range: 17.5–18.5 mo) and 36 36-mo-olds (TB-TB condition: 18; TB-FB condition: 18; mean age, 36.3 mo, range: 35.0–36.9 mo).

Materials. A toy egg and a toy carrot were used in the warm-up trials. Two novel objects made specifically for this study were used in the test trials. Two cardboard boxes (a green one and an orange one) with a lid were used as hiding containers. A pair of ordinary (transparent) sunglasses was used in the familiarization phase, and different but similar-looking sunglasses were used in the test phase, which were transparent in the TB-TB condition and opaque in the TB-FB condition.

Procedure. The procedure was a modified version of the task of Southgate et al. (18).

Familiarization. Children were seated on the floor with their parent and were shown a pair of ordinary (transparent) sunglasses by E1 to make sure that they were familiar with the object and its use.

Warm-up trials. E1, wearing sunglasses on her head as a hairband, kneeled in between the two cardboard boxes, which were 100 cm apart and 120 cm from the child. First she gave the child a toy egg and a toy carrot to play with for roughly 10 s. She then placed one object in each box and asked the child to retrieve first one then the other (by naming them). The hiding game continued until the child correctly chose the requested objects twice in a row from two different boxes.

Test trial. The test trial consisted of three phases: a belief induction phase, a belief revision phase, and a test phase.

Belief induction phase. E1 gave the children the two novel objects to explore for about 10 s. These objects were not labeled in this phase. E1 then placed one object in each box and closed the lids. The location of the objects was counterbalanced across infants. At this point, E2 asked E1 to put on her sunglasses. E1 then put her sunglasses from her head on her eyes and sat back, but stayed in the room facing the subsequent events. E2 then deceptively approached the boxes (gesturing "shush" toward the child, following the protocol of ref. 18), switched the objects, closed the boxes, and asked E1 to leave the room with her. Before leaving, E1 removed her sunglasses and left them in front of the participant.

Belief revision phase. At this point, the child was encouraged by the parent to try on the sunglasses. Before the experiment, the parents had been told that, when they are left alone, they should ask their child to try on the sunglasses and verify together whether they could see through them. The parents had been informed in advance whether the sunglasses were opaque or transparent to avoid explicit signals of their own surprise. Importantly, in the TB-TB condition, the sunglasses were transparent, but in the TB-FB condition, they were opaque.

Test phase. After being away for ~45 s, E1 returned to the room, greeted the infant, and sat on the floor behind the two boxes. E1 then pointed at one of the boxes (counterbalanced across infants) and said (in Hungarian), "Do you remember what I put here? I put a sefo here. Shall we play with the sefo?" alternating gaze between the infant and the referred box twice ("sefo" is a phonotactically valid pseudoword in Hungarian). E1 then grasped both boxes, extended her arms toward the child, and simultaneously opened the lids of both boxes that were oriented toward the child while looking at the child. At this point, the contents of the boxes became visible only to the child. E1 then said, "Can you give me the sefo?" while looking directly at the child and not looking toward either box. E1 repeated the question until the child began to approach one of the boxes or pointed toward one of the boxes, or until 180 s had passed.

Coding. The sessions were video-recorded and coded offline. All parents followed the instructions and ensured that, during the belief revision phase, their children noticed whether they could see through the sunglasses. The parental utterances during this phase were transcribed and analyzed. The analyses yielded no significant difference in the number or in the length of the utterances across the two age groups (for details, see the *SI Appendix*).

The dependent measure during the test phase was the choice that children made in response to E1's request. The first response toward one of the boxes, after E1 had said, "Can you give me the sefo?" was coded as the child's choice and was categorized as choosing the referred or the nonreferred box. Both reaching and pointing responses were accepted as valid choices. All sessions were coded also by a second observer, blind to the experimental condition, watching only the recordings of the test phase. Interrater agreement was 96% (Cohen's Kappa, 0.91).

Experiment 2.

Participants. Twenty 18-mo-old infants were recruited for this experiment. Because of experimenter error, one infant was excluded and was replaced. Of the 20 infants, five did not make a choice during the test phase. The remaining 15 infants (mean age, 18.0 mo; range, 17.5–18.5 mo) constituted the final sample for this experiment.

Materials. The same props were used as in Experiment 1, with the exception of the sunglasses. A single pair of opaque sunglasses was used in this study.

Procedure. The procedure used in this experiment was identical to Experiment 1, with the following exceptions. Infants were shown a pair of opaque sunglasses to explore during familiarization, and E1 wore the very same sunglasses during the belief induction phase (during the location change). Then E1 left the room for about 45 s, leaving the pair of sunglasses behind. However, parents were not instructed to explore the sunglasses with their child. Thereby the belief revision phase was skipped but the delay between the belief induction and test phases was kept identical to that of Experiment 1.

Coding. The coding of the dependent variable was the same as in Experiment 1. Interrater agreement was 93% (Cohen's Kappa, 0.87).

Experiment 3.

Participants. Forty 36-mo-old children were recruited. Three children were excluded because of experimenter error, and were replaced. Of the 40 infants, four did not make a choice during the test phase. The remaining 36 infants (TB-TB condition: 18; TB-FB condition: 18; mean age, 36.3 mo; range, 35.1–36.9 mo) constituted the final sample.

Materials. The same props were used as in Experiment 1, with the exception of the sunglasses.

Procedure. The procedure was similar to Experiments 1 and 2, with the exception that the belief manipulation was not implemented by sunglasses, but by an unexpected one-way mirror. The warm-up trials were similar to Experiment 1.

Belief induction phase. This was identical to Experiment 1 and 2, up to the point when the objects were placed in the boxes. At this point, E2 asked E1 to leave the room. E2 then approached the boxes, switched the location of the objects, and closed the boxes.

Belief revision phase. The child was asked by E2 to call E1 back from the adjacent room (parents were allowed to join). They entered the other room, where children saw E1 either peeking into the experimental room through a one-way mirror (FB-TB condition) or reading a book while the mirror was covered (FB-FB condition). In the FB-TB condition, E2 encouraged the child to look through the one-way mirror. After ∼45 s, E2 asked E1 and the child to return to the experimental room.

Test phase. This was identical to that of Experiment 1.

Coding. Coding of the dependent variable was the same as in Experiment 1 and 2. Interrater agreement was 97% (Cohen's Kappa, 0.944).

ACKNOWLEDGMENTS. We thank parents and children participating in this study and D. Árva, J. Baross, I. Savos, and Zs Kalina for their help with data collection; R. Schvajda for the transcripts; and F. Elekes and D. Kampis for their valuable comments on the manuscript. This study was supported by Grant OTKA 116 779 (to I.K.). Partial funding came from European Union's Seventh Framework Programme (FP7/2007-2013), ERC Grants 609819 SOMICS, and 284236 REPCOLLAB.

1. Boyd R, Richerson P-J (1996) Why culture is common, but cultural evolution is rare. *Proc Br Acad* 88:77–93.
2. Baron-Cohen S, Leslie A-M, Frith U (1985) Does the autistic child have a "theory of mind"? *Cognition* 21:37–46.
3. Wimmer H, Perner J (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13:103–128.
4. Onishi KH, Baillargeon R (2005) Do 15-month-old infants understand false beliefs? *Science* 308:255–258.
5. Surian L, Caldi S, Sperber D (2007) Attribution of beliefs by 13-month-old infants. *Psychol Sci* 18:580–586.
6. Kovács Á-M, Téglás E, Endress AD (2010) The social sense: Susceptibility to others' beliefs in human infants and adults. *Science* 330:1830–1834.
7. Southgate V, Senju A, Csibra G (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci* 18:587–592.
8. Rubio-Fernández P (2013) Perspective tracking in progress: Do not disturb. *Cognition* 129:264–272.
9. Buttelmann D, Carpenter M, Tomasello M (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112:337–342.
10. Knudsen B, Liszkowski U (2012) 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy* 17:672–691.
11. Kovács Á-M (2016) Belief files in theory of mind reasoning. *Rev Phil Psychol* 7:509–527.
12. Ganea PA, Harris PL (2010) Not doing what you are told: Early perseverative errors in updating mental representations via language. *Child Dev* 81:457–463.
13. Song HJ, Onishi K-H, Baillargeon R, Fisher C (2008) Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition* 109:295–315.
14. Tauzin T, Gergely G (2018) Communicative mind-reading in preverbal infants. *Sci Rep* 8:9534.
15. Klein S-B, et al. (2009) Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Soc Cogn* 27:283–319.
16. Nelson K, Fivush R (2004) The emergence of autobiographical memory: A social cultural developmental theory. *Psychol Rev* 111:486–511.
17. Hayne H (2004) Infant memory development: Implications for childhood amnesia. *Dev Rev* 24:33–73.
18. Southgate V, Chevallier C, Csibra G (2010) Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Dev Sci* 13:907–912.
19. Powell L-J, Hobbs K, Bardis A, Carey S, Saxe R (2017) Replications of implicit theory of mind tasks with varying representational demands. *Cognit Dev* 46:40–50.
20. Eacott M-J, Crawley R-A (1998) The offset of childhood amnesia: Memory for events that occurred before age 3. *J Exp Psychol Gen* 127:22–33.
21. Scarf D, Gross J, Colombo M, Hayne H (2013) To have and to hold: Episodic memory in 3- and 4-year-old children. *Dev Psychobiol* 55:125–132.
22. Senju A, Southgate V, Snape C, Leonard M, Csibra G (2011) Do 18-month-olds really attribute mental states to others? A critical test. *Psychol Sci* 22:878–880.
23. Sperber D, Wilson D (2002) Pragmatics, modularity, and mind-reading. *Mind Lang* 17: 3–23.
24. Setoh P, Scott RM, Baillargeon R (2016) Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proc Natl Acad Sci USA* 113: 13360–13365.
25. Gerstenberg T, Bechlivanidis C, Lagnado DA (2013) Back on track: Backtracking in counterfactual reasoning. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds Knauff M, Pauen M, Sebanz N, Wachsmuth I, (Cognitive Science Society, Austin, TX), pp 2386–2391.
26. Schacter D-L, Guerin S-A, St Jacques PL (2011) Memory distortion: An adaptive perspective. *Trends Cogn Sci* 15:467–474.
27. Cochran K-J, Greenspan R-L, Bogart D-F, Loftus E-F (2016) Memory blindness: Altered memory reports lead to distortion in eyewitness memory. *Mem Cognit* 44:717–726.
28. Bauer P-J, Leventon J-S (2013) Memory for one-time experiences in the second year of life: Implications for the status of episodic memory. *Infancy* 18:755–781.
29. Bauer P-J, Wenner J-A, Dropik P-L, Wewerka S-S (2000) *Parameters of Remembering and Forgetting in the Transition from Infancy to Early Childhood* (Wiley, New York).
30. Mullally SL, Maguire E-A (2014) Learning to remember: The early ontogeny of episodic memory. *Dev Cogn Neurosci* 9:12–29.
31. Martin A, Santos LR (2016) What cognitive representations support primate theory of mind? *Trends Cogn Sci* 20:375–382.
32. Butterfill S, Apperly I (2013) How to construct a minimal theory of mind. *Mind Lang* 28:606–637.
33. Perner J, Ruffman T (2005) Psychology. Infants' insight into the mind: How deep? *Science* 308:214–216.
34. Atance C-M, Sommerville J-A (2014) Assessing the role of memory in preschoolers' performance on episodic foresight tasks. *Memory* 22:118–128.
35. Suddendorf T, Nielsen M, von Gehlen R (2011) Children's capacity to remember a novel problem and to secure its future solution. *Dev Sci* 14:26–33.
36. Tulving E (2005) Episodic memory and autonoesis: Uniquely human. *The Missing Link in Cognition*, eds Terrace H, Metcalfe J (Oxford Univ Press, New York), pp 4–56.