Infants and adults neurally represent the perspective of others like their own perception

Authors: Anna-Lena Tebbe[1], Katrin Rothmaler[1], Moritz Köster[2], Charlotte Grosse Wiesmann[1]*

Affiliations:

[1]Milestones of Early Cognitive Development, Max Planck Institute for Human Cognitive and Brain Sciences; Leipzig, 04103, Germany.

[2]Institute for Psychology, Developmental Cognitive Psychology, University of Regensburg; Regensburg, 93055, Germany.

*Corresponding author. Email: wiesmann@cbs.mpg.de

Abstract:

Preverbal infants already seem to consider the perspective of others, even when it differs from their own. Similarly, adults take the perspective of others very quickly, in parallel to other cognitively demanding tasks. This raises the question of how multiple perspectives are processed efficiently, and even before higher cognitive capacities develop. To test whether and how others' perspectives are neurally represented, we presented 12-14-months-old infants and adults with objects flickering at 4 Hz, which evoked neural oscillations at the exact same frequency. Remarkably, both in infants and adults, this same highly specific neural signature of visual object processing was also present when their view was blocked and only another observer saw the object. These results provide strong evidence that we process what others see as if we saw it ourselves, revealing a neural mechanism for efficient perspective taking, present from infancy.

Introduction

As highly social beings we are continually confronted with the difficult task of considering others' perspectives. Communicating with others, interpreting their behavior, deciding about our own next steps; human thought and interaction is hardly imaginable without perspective taking. Yet, how this capacity is achieved, and how it develops, has puzzled empirical researchers and theorists. To address this, we devised an experimental approach that allowed us to test how the content of other people's perspective is represented neurally, in adults as well as in preverbal infants.

Traditionally, reasoning about others' mental representations is thought to emerge late in the preschool years and to rely on higher cognitive functions (i.e., cognitive control and language), as several different representations of the world must be maintained in parallel – the others' versus one's own perspective (*1*, *2*). However, this view has been challenged by evidence that, in their behavior, preverbal infants already seem to consider other people's mental states, including their visual perspective (*3*, *4*), preferences and intentions (*5*), knowledge (*6–8*), and possibly beliefs (*9*, *10*), even when they differ from the infant's own. These findings have created a puzzle: If infants actually do represent others' mental states, how do they manage multiple parallel representations of their environment before higher cognitive capacities develop?

In addition to the capacities observed in infancy, numerous studies in adults suggest there is an inbuilt and computationally efficient way of considering others' mental representations. For example, we consider others' mental states in real time, as we speak or act, very quickly and online, in parallel to other tasks (e.g., *11–14*). Adults even seem to compute other people's mental states when they are irrelevant to their current task (e.g., *15*, *16*). For example, they recognize or detect an object faster when another person is looking at it (*17*) or falsely believes it to be in a location (*16*). Similarly, infants' expectations and processing of their environment are influenced by what others see or believe (*15*, *16*, *18*, *19*). These findings suggest that both infants and adults compute others' mental representations automatically, even when dedicating their attention to something else, and so quickly that it can influence perceptual tasks such as object detection.

The past decade has seen an intense theoretical debate in an attempt to explain how perspective taking could be so fast and efficient despite the computational demands of mastering different representations. A number of researchers have suggested that infants do not, in fact, represent others' mental representations at all. They argue that the presence or gaze of the others may simply enhance the infants' attention (*20*, *21*) or boost their memory (*22*). Similarly, it has been questioned whether adults actually represent others' mental states in such situations (*21*, *23*). Contrary to traditional assumptions, other researchers, have proposed that perspective taking is *not* cognitively effortful, but that an efficient format exists that is present from infancy (*16*). However, this leaves open the question of what this efficient format could be.

To address these questions, we designed an experiment targeting infants' and adults' neural representations of objects they see themselves. We then tested whether and how they neurally represent those same objects when the objects are exclusively seen by another person and not themselves. We hypothesized that, rather than inferring others' perspectives separately from their own, adults and infants represent the objects others see in the same format as their own perception, that is, *as if they saw the objects themselves*. Thus, participants should show a neural signature specific to their own visual processing of an object, not only when seeing the object themselves, but also when observing another person seeing the object. Representing what others see, in the format of one's own perception, would equate to a means of processing this content as quickly and easily as seeing it oneself. Further, because such a mechanism would require similar cognitive demands to those of perception, we predicted that such a process could quite plausibly already operate in infancy.

To test this, we presented adults (experiment 1) and 12- to 14-month-old infants (experiment 2) with videos featuring an agent and an object. The agent watched the object move behind an occluder (so the agent could still see it while the participant's view was blocked, Fig. 1). In a control condition, the object moved into a tunnel (blocking both the agent's and the participant's view). The object was flickered on and off, in a constant rhythm of 4 times per second, within the static, non-flickering scene (Video S1). Prior studies have shown that the specific frequency of a visually flickered image is faithfully replicated by rhythmic brain activity over the visual cortex at the exact same frequency and phase (*24*, *25*). Importantly, such phase-locked rhythmic activity does not occur spontaneously in absence of a flickering stimulus (*26*). Thus, participants'

phase-locked 4 Hz response provided a highly specific neural signature of their visual processing of that particular object. Based on our hypothesis, we predicted the following: Participants' visual cortices should show this evoked rhythmic response, not only while seeing the object themselves, but also after object occlusion, in trials where the agent can still see the object. In contrast, when the agent's view is blocked by the tunnel, the evoked 4 Hz activity should be absent.

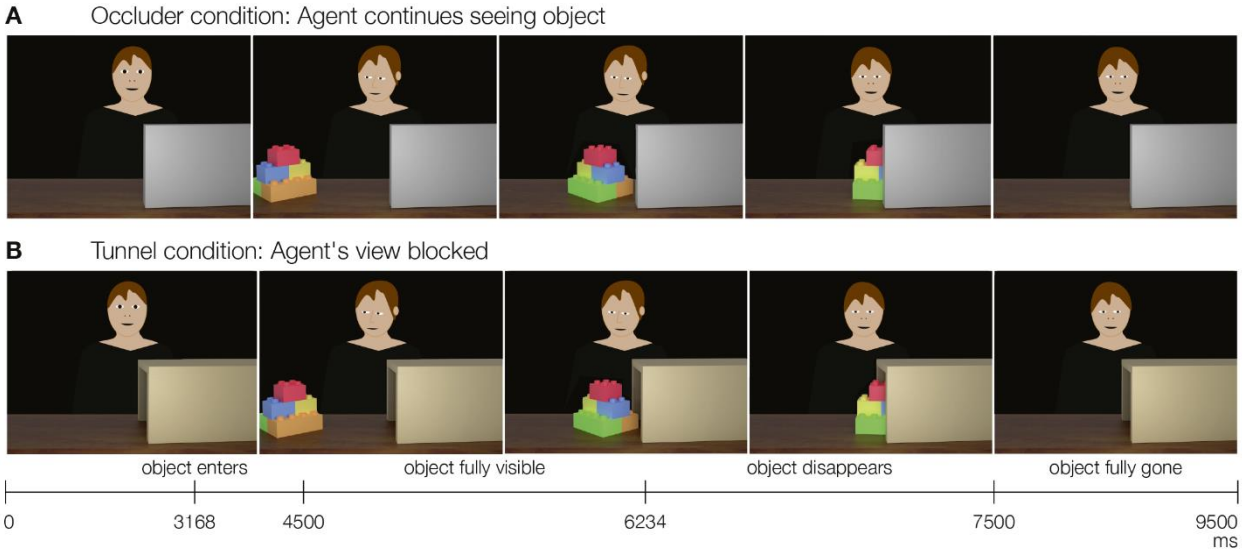

Fig. 1. Schematic illustration of the time course of two exemplary trials. (**A**) Flickering objects either moved behind an occluder (continued visual access for the agent only) or (**B**) into a tunnel (neither participant nor agent have visual access). Please note that, due to image rights, the real person shown in the videos has been replaced by a schematic figure.

Results

To determine the amplitude of participants' evoked 4 Hz response at every time point during the trial, we computed the envelope of the phase-locked 4 Hz oscillations over the visual cortex using the Hilbert transformation in adults. Because infants contributed an unbalanced number of trials per condition (see Methods), we computed the signal-to-noise ratio (SNR) as a baseline-free measure of their evoked 4 Hz signal (for comparison, the SNR was also calculated in adults yielding similar results, see S3.1). Both adults' (N = 40) and infants' (N = 56) evoked 4 Hz signal increased as a function of the object's visibility to the participant (adults: Fig. 2; infants: Fig. 3). This established that viewing the flickering object, moving through the static scene, indeed evoked phase-locked rhythmic activity at 4 Hz, located over occipital sites (Fig. 2B, Fig. 3B).

Further, both samples' frequency spectra showed clear peaks at 4 Hz and its harmonics, as expected for steady state visually evoked potentials (SSVEPs) (Fig. 2A, Fig. 3A).

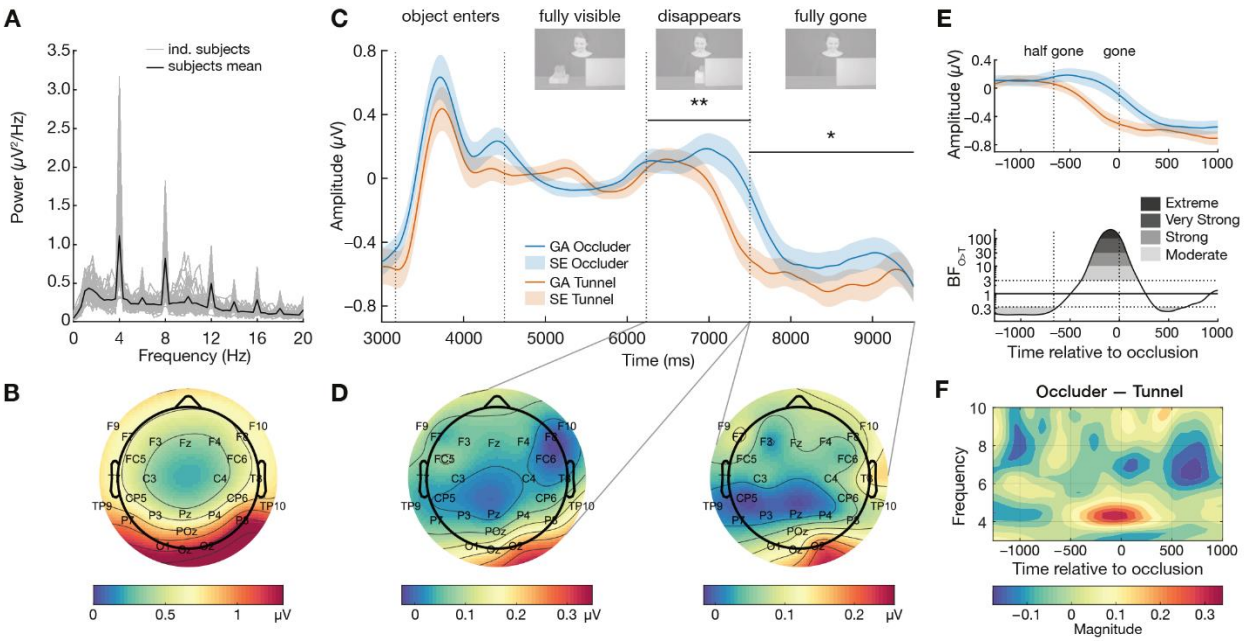

Fig. 2. Results of experiment 1: Evoked oscillations at 4 Hz in *N* = 40 adults. (**A**) The frequency spectrum shows clear peaks at 4 Hz and its harmonics (electrode Oz). (**B**) The topographical results demonstrate a clear SSVEP to the flickering object over visual cortex. (**C**) Baseline-corrected average evoked 4 Hz signal envelope over occipital sensors (O1, O2, Oz). Results show a higher 4 Hz response when the agent continued to see the object in the Occluder condition (blue) compared to the Tunnel condition (orange), both in the disappearance phase (strong evidence, indicated by two stars) and occlusion phase (moderate evidence, one star). Shading indicates the standard error. (**D**) Topography of the evoked 4 Hz Occluder-Tunnel difference while the object disappears (left) and during full occlusion (right). (**E**) Evoked 4 Hz amplitude envelope in the period from object disappearance until 1000 ms after full occlusion (top, time relative to object occlusion) and BF_{10} of the dynamic comparison between conditions (bottom) showing higher 4 Hz response in Occluder compared to Tunnel trials (BF_{10} between 3–424.45). Note that the y-axis of the BF_{10} is log-scaled. (**F**) The time-frequency plot displays the power difference between the Occluder and Tunnel conditions at the end of the trial, with a clear 4 Hz peak before and after the object was fully occluded.

Experiment 1: Adults process what others see like their own visual percepts

Our central prediction was that this evoked 4 Hz signal would also be present after object occlusion when the agent continued to see the object (Occluder condition), as opposed to when her view was blocked by the tunnel (Tunnel condition). Indeed, in adults there was extremely

strong evidence for a higher evoked 4 Hz response in the Occluder condition compared to the Tunnel condition, specifically during the time window when the object disappeared, and lasting until after the object was fully occluded (dynamic Bayesian t-tests: BF_{10} ranging between 3 and 424.45, from 7050 ms to 7736 ms, see Fig. 2C & 2E). These effects remained when the evoked 4 Hz signal was averaged within two time intervals: while the object was fully occluded (Bayesian Wilcoxon signed-rank test: $BF_{10} = 6.16$, W = 568, $\hat{R} = 1.011$), and while the object was disappearing ($BF_{10} = 20.02$, W = 601, $\hat{R} = 1.023$). Adults' visually evoked response to the flickering object was thus prolonged when the agent continued to see the object and remained present, despite the fact that the object was no longer visible to the participants themselves. Hence, instead of two distinct representations - one for their own perception of the environment and one for the others' representation - adults processed what the other person saw similar to seeing the flickering object themselves. This perceptual format offers a way of considering the content of the others' mind as quickly and easily as our own perception.

Experiment 2: Infants process what others see like their own visual percepts

Reducing the cognitive demands of perspective taking to those of perception may also provide infants with a solution to this task, despite their limited cognitive capacities. In experiment 2, we therefore presented 12- to 14-month-old infants with the same videos and compared their evoked 4 Hz signal in the Occluder versus the Tunnel condition. As in adults, this yielded very strong evidence for higher evoked 4 Hz responses in the Occluder condition compared to the Tunnel condition. In infants, this occurred in a time window after the object was fully occluded (dynamic Bayesian t-tests: BF_{10} ranging from 3 to 74.45 between 8460 ms and 8752 ms, Fig. 3). Averaged across the entire disappearance and occlusion phase, no evidence for a difference was found (Occlusion phase: Bayesian Wilcoxon signed-rank $BF_{10} = .66$, W = 629, $\hat{R} = 1.005$; Disappearance phase: $BF_{10} = .18$, W = 509, $\hat{R} = 1.000$). This suggests that, for infants, the effect was specific to the observed time window after object occlusion.

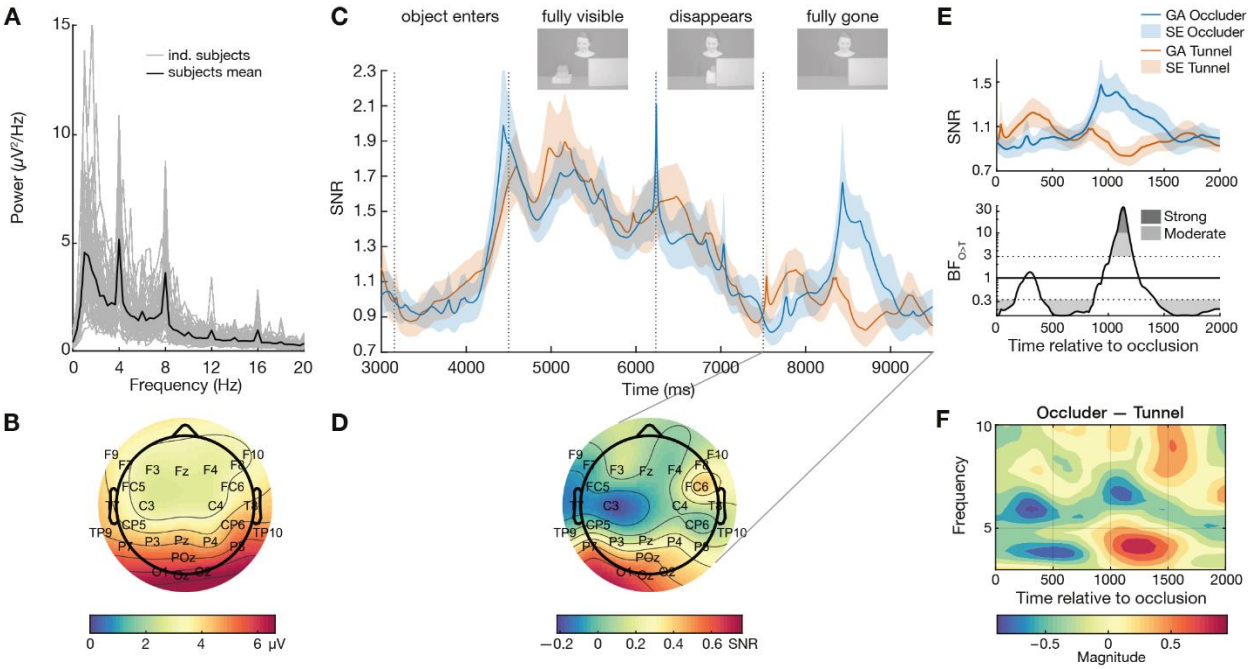

Fig. 3. Results of experiment 2: Evoked oscillations at 4 Hz in *N* = 56 infants. (**A**) The frequency spectrum shows clear 4 Hz peaks and harmonics (electrode Oz). (**B**) The topographical results show a clear SSVEP to the flickering object over visual cortex. (**C**) Average 4 Hz SNR over occipital sensors (O1, O2, Oz). Results show a higher evoked 4 Hz response when the agent sees the object in the Occluder condition (blue) compared to the Tunnel condition (orange) in a time period after object occlusion. Shading indicates the standard error. (**D**) Topography of the evoked 4 Hz Occluder-Tunnel difference during full occlusion. (**E**) Evoked 4 Hz signal in the time window from object disappearance until the end of the trial (top, time relative to object occlusion) and the BF_{10} of the dynamic comparison between conditions (bottom) showing a higher 4 Hz response in Occluder compared to Tunnel trials (BF_{10} between 3-74.45). (**F**) The time-frequency plot displays the power difference between the Occluder and Tunnel conditions at the end of the trial with a clear 4 Hz peak after the object was fully occluded.

Thus, similar to adults, infants showed an evoked 4 Hz response to the flickering object, even though the object was no longer visible to them, but only remained visible to the other person. However, infants only showed this effect in the time window after the object was fully occluded. In adults, the effect was already present while the object was in the process of disappearing, and then lasted throughout the object occlusion. Such differences in latency are common between adult and infant processing. Quite possibly, adults simulate the events ahead of time, anticipating the differences in the agent's visual access before the object is fully occluded. Infants, in turn, may have needed more time to infer the agents' visual access to the object, and their realization of it then seemed to reactivate their own visual representation of the flickering object, triggering visual processing of the object although fully occluded.

Control analyses

Although the visual features of the occluder and tunnel (e.g., surface area, luminance) were closely matched and their colors counterbalanced, the minimal visual differences required to show an occluder versus a tunnel could have contributed to the difference in evoked 4 Hz signal, in particular, during the object disappearance. To rule out this possibility, we additionally acquired a non-social control condition where the object moved behind the occluder, but no agent was present watching the object. When no agent was present, the evoked 4 Hz signal was neither prolonged nor enhanced compared to the tunnel condition (see S3.3). This reinforces that the observed difference indeed resulted from the agent's visual access to the object and not from visual differences between the tunnel and the occluder. Further, no difference between conditions was found over frontal electrodes, including when eye movements were retained in the signal (see S3.4). This confirms that the observed difference between the Tunnel and Occluder conditions did not result from differential eye movements between the conditions, which would have been visible in frontal electrodes.

Discussion

Together, our findings demonstrate that both adults and infants show a highly specific neural signature of their own visual processing of an object when observing another agent seeing the object, even if their own view of the object is blocked. This is not the case when the agent's view is also blocked, or no agent is present. This provides direct evidence that infants and adults process what only others can see like seeing these objects themselves. The finding that their visual object processing was sustained and even reactivated after full object occlusion suggests that, when seen by an agent, objects are kept online, or reactivated, in visual working memory, where they are processed similar to perception. This is corroborated by the specificity of this effect to 4 Hz (see S3.5, S3.6), which, as a spontaneous intrinsic rhythm in the brain, supports working memory (*27–29*). It is important to note that this effect cannot reflect a general increase in working memory load, as this would have led to an increase in spontaneous 4 Hz oscillations at arbitrary phases, which would have canceled each other out and not been visible in the reported *phase-locked* evoked 4 Hz signal. Such evoked responses do not occur spontaneously (without rhythmic visual stimulation) and do not typically persist after the flickering stimulus is no longer visible (*26*). The observed response is thus highly specific to sustaining one's own *visual* processing of the object. Hence, it provides clear evidence that infants and adults represent

what others see by activating their own visual representations of the object. When observing the agent seeing the flickering object, the object seems to continue flickering in our mind's eye.

An efficient neural mechanism for considering other perspectives

Processing what others' see, as if seeing it oneself, provides a mechanism for how adults and infants can consider others' perspectives as easily as their own perception, and in a similar time scale. They do not need to compute and maintain distinct representations for their own and the others' perspective. Instead, it is sufficient to infer the agent's visual access and then simply reactivate their own visual percepts of objects, seen by another person. This explains why adults can do this quickly and under high cognitive load (e.g., while speaking (*13*) or acting (*14*, *30*)). Moreover, it explains how even infants, despite limited processing capacities, can consider perspectives that differ from their own (*2*, *3*, *9*, *10*) and why others' visual access can modulate their neural processing of a scene (18). Processing others' perspectives in a perceptual format also provides an explanation for why this perspective could interfere with our own perception and thus modulate our actions and expectations of the environment. This has been observed behaviorally in both adults (*16*, *17*, *31*, *32*) and infants (*15*, *16*, *19*). Importantly, previously used behavioral and neural indicators were not specific to perception, leaving open the question of the neural format that allows for such modulations. Based on findings that adults recognized rotated symbols faster when another person saw these upright, it has been proposed that we process what others see in a quasi-perceptual format (*17*). Confirming this proposal, we show that objects seen by others indeed activate individual perceptual processes, in adults and in infants, even when not visible to the participants themselves.

It is noteworthy that previous explanations of behavioral findings according to which following the others' line of sight simply enhances attention to the object, in the form of a directional cue (*20*, *23*, *33*), are ruled out by our design. First, the object is no longer visible to the participant so their attention to it cannot have been enhanced. Second, the agent's gaze direction was identical between experimental conditions, and the agent attended the object equally before its occlusion. This also speaks against a recent account that, when infants attend to an object together with an agent, this leads to better encoding of the object while it is visible (*22*). Instead, in our paradigm, infants need to infer the agent's visual access and our findings show that they then actually *represent* what the other sees. However, they do so by reactivating their own visual percept of

the object (i.e., like seeing it themselves or holding it in visual working memory) and not as a separate representation. This way, the object is readily available to their own processing of the environment. As this perceptual representation is maintained, despite not seeing the object oneself, it is likely also available for other mental states than the visual perspective. For example, our visual working memory of an object in a certain location may be enhanced in the same way, if an agent has seen the object in a certain location in the past and thus believes it to be there. This would explain why modulations of adults' and infants' behavior on the environment have also been found when agents had a false belief (*14–16, 19*).

An exciting question for future research is how this perceptual processing of others' perspective is related to reasoning about others' mental representations *separately* from our own representation of the environment. There is research indicating that perspective taking processes present from infancy may be different from those that support verbal reasoning about mental states *as distinct* from our own (*34, 35*). The latter seem to rely on passing visual representations on to higher cognitive networks (e.g., *38, 39*), and require an exchange with brain structures involved in language and executive function, maturing later in childhood (*38–40*).

However, young infants lack such higher cognitive capacities but are strongly dependent on others. Not only do they rely on others to fulfill their basic needs, but others guide them in making sense of, and learning from, a highly complex environment. Having access to what others attend to, or hold in mind, even when infants cannot see it themselves, is essential to predicting caregivers' actions, to following the targets of their attention, and to understanding the referent of their communication. Processing this perceptually, within their own representation of the world, provides a mechanism for infants to understand others and learn from them without the cognitive demands of managing multiple different representations. And not only infants. This mechanism persists and appears to become refined in adulthood, offering a fast and highly efficient way of considering what others hold in mind. We do not need to infer the perspectives of others around us as a perspective distinct from our own, but, in lack of resources, seem to *see what the other sees in front of our mind's eye*. This intuitive understanding of, and alignment with, others, from very early on in life, may be a driver of human cooperation and at the basis of our unique capacity to learn from and teach others (*41, 42*). As such, the revealed mechanism

may constitute an adaptive advantage leading to the ultra-social nature of human cognition, forming a foundation for the evolution of human culture (*43*, *44*).

Methods

The video stimuli, experimental procedure and data processing was identical in Experiment 1 (adults) and Experiment 2 (infants) unless otherwise specified.

Experiment 1 (adults)

Sampling approach

Our sample size was based on a preregistered Bayesian sequential testing scheme (*45–48*), i.e., data collection was continued until the Bayes factor (BF) of our main hypothesis (modulation of the sustained steady state signal depending on social context during the disappearance phase) reached 3 or 1/3. We tested a minimum sample size of $N = 30$ (*49*) and continued testing until sufficient statistical evidence for or against a difference between the experimental conditions was reached, yielding a sample size of $N = 31$. By the time of final analyses, nine additional adult participants had been collected so that the reported data include $N = 40$ participants (highly similar results for $N = 31$ see suppl. S6).

Participants

Experiment 1 reports data of $N = 40$ healthy adult participants (15 female, 25 male; $Mean_{age}$ = 26.05 years, $SD = 3.66$). Two additional participants were excluded from the final sample due to poor EEG data quality (more than 50% EEG trials with artifacts). None of the participants reported any history of neurological or psychiatric disorders or photic epilepsy. All participants had normal or corrected to normal vision,provided written informed consent prior to participation and were compensated financially for their time. The study received approval from the Ethics Committee of the University of Leipzig.

Stimuli and Procedure

In a within-subject design, we presented participants with videos, in which an agent observed a flickering object moving across a table. Towards the end of each trial, the object disappeared behind an occluder, or in a control condition into a tunnel. Critically, in the Occluder condition, the agent continued to see the object, whereas in the Tunnel condition, her view was blocked.

Each trial started with an attention getter presented for 1000 ms (animal picture and sound), followed by a black screen, presented for a random duration between 500 and 800 ms. Subsequently, one of the videos with the flickering objects was shown. The videos featured an actress positioned at the center of the scene. The agent smiled briefly, then gazed towards the left bottom of the table, where an object appeared shortly after. The object moved across the table, followed by the gaze of the agent and was fully visible for 1733 ms, defining the *object phase*. The object then took 1776 ms to move behind the Occluder or inside the Tunnel at the right end of the table (defining the *disappearance phase*). After this, the object was completely occluded for 2000 ms (*occlusion phase*). The duration of the videos was 9.5 sec. In an additional control condition, the object disappeared behind the occluder, but no agent was present watching it (see S2).

Participants were presented with 60 videos per condition, organized in blocks of 30 videos of the same condition. The order of the blocks was counterbalanced across participants. Trials featured 20 different objects that were shown in random order. To make the conditions easily distinguishable, Occluder and Tunnel were shown in two different colors (brown and gray) that were matched for luminance and counterbalanced across participants. In order to stress the different spatial properties of Occluder versus Tunnel, and to ensure that participants understood the agent's visible access to the object depending on condition, each of the blocks was preceded by a familiarization video. These showed how the Occluder or Tunnel (depending on condition), entered the scene, made an 180° turn, and then moved into its final position, while the agent followed it with their gaze.

Videos were created using the graphics software Blender (v. 2.8, www.blender.org). The stimuli were presented using Psychtoolbox (version: 3.0.17; *50*, 5*1*) in Matlab (R2017a; Mathworks, Inc., Natick, MA, USA) and were played with 60 frames per second (i.e., 16.67 ms refresh interval) at the center of a 17-inch CRT screen at a visual angle of 12° x 22°. To establish a 4 Hz flicker, we presented the frames of our videos at a duty cycle of 7:8 (i.e., seven frames (116.69.ms) showing a bright object ('on'), followed by eight frames showing the same object strongly dimmed ('off'), followed by 8 on-frames and 7 off-frames (*52*). Importantly, not the whole scene but only the object was presented in a "flickering" mode. Participants viewed the display at a distance of 110 cm and were instructed to watch the videos attentively and to avoid

eye movement and blinks during video presentation. To maintain participants' attention throughout the experiment, they were asked control questions on properties of the object (e.g., color, shape, appearance) after randomly selected trials (two questions within every 10 trials). The questions could be answered with 'yes' or 'no' by pressing one of two mouse buttons.

EEG Recordings and Data Preprocessing

EEG was recorded continuously using 28 active electrodes (FC5/6, Cz, C3/4, T7/8, CP5/6, TP9/10, Pz, P3/4, P7/8, Poz, Oz, O1/2, ground electrode located at FP1, placed according to the International 10–20 System and secured in an elastic electrode cap (Easycap GmbH, Herrsching, Germany). Electrode impedances were kept below 30 kΩ. Data were digitalized with a sampling rate of 500 Hz with Cz as reference and processed offline using Brain Vision Analyzer (Version 2.1.2.327) and MATLAB (2023a; Mathworks, Inc., Natick, MA, USA). We excluded epochs with major artifacts by visual inspection. We further excluded segments with significant artifacts identified via automatic artefact correction (1) gradient (maximal allowed voltage step): 50 µV/ms; 2) difference in amplitude (maximum-minimum): 300 µV in an interval of 300 ms; 3) minimum/maximum allowed amplitude of ±500 µV; 4) lowest allowed activity in intervals of 100 ms: 0.5 µV). Data was band-pass filtered to 1-20 Hz using an IIR Butterworth filter with zero phase shift. Eye Movements were identified and corrected automatically using the Ocular Correction Independent Component Analysis (ICA) integrated in the Brain Vision Analyzer Software and discarded after visual inspection. EEG epochs were segmented from 0 to 9500 ms post-onset.

SSVEP Analysis

Artifact-free epochs were averaged separately for each condition. Data were bandpass filtered with a 10th order zero phase Butterworth filter having a halfwidth of 1 Hz around the stimulation frequency. To avoid filter edge artifacts, value padding was performed at the beginning and the end of the signals (500 values on each side). The time-varying amplitude of the SSVEP response was extracted via the Hilbert transform. For the statistical analysis, a cluster average was computed for the region of interest (O1, O2, Oz) based on the expected topography and main cortical source of the SSVEP signal (*53*, *54*). To eliminate offset differences between conditions, the mean amplitude of the object phase (i.e., the time window during which the object was completely visible) was subtracted as baseline.

Statistical analyses

Statistical analyses were carried out with Matlab (R2023a) and JASP (*55*). We performed Bayesian analyses and used default priors for all models, i.e., an equal, uninformed prior probability as we had no a priori information regarding the effects. Following conventions, we considered a BF_{10} between 3 and 10 as *moderate* evidence, between 10 and 30 as *strong* evidence, between 30 and 100 as *very strong* evidence, and larger 100 as *extreme* evidence for the given hypothesis H1 (*56*). Analogously, a BF_{10} between 1/3 and 1/10 was considered *moderate* evidence, between 1/10 and 1/30 as *strong* evidence, between 1/30 and 1/100 as *very strong* evidence, and smaller than 1/100 as *extreme* evidence in favor of H0. BFs between 1/3 and 3 were considered inconclusive.

The signal envelopes were analyzed dynamically by calculating Bayesian directed paired t-tests between the conditions at each time point from when the object started disappearing until the end of the trial (*26, 27*). In addition, the signal envelopes were averaged across two time windows: 1) the *object disappearance phase*, starting from the first time point at which the objects start to disappear until 80 frames [~1333 ms] later, and 2) the *occlusion phase*, starting from the time point at which the object is completely occluded until the end of the trial. As q-q-plots (see S7) indicated minor deviations from normality, we used we used non-parametric statistical test alternatives whenever indicated and available.

Experiment 2 (Infants)

Participants

Similar to the adults, our final infant data set resulted from a Bayesian sequential testing scheme with a minimum N of 40 and stop criterion of BF = 3 or 1/3 (*48*), leading to a sample of $N = 56$ healthy full-term infant participants aged between 12 and 15 months (31 female; $Mean_{age} = 13.4$ months, $SD = 22.6$ days). Fifteen additional subjects were excluded from the sample because they did not provide at least five artifact-free trials per condition. Due to the manual coding of infant's looking behavior and the complex analyses, we recorded an additional 16 infants after the stopping criterium was met. In the results, we report the full sample, but the results at the stopping criterion $N = 40$ yields highly similar results (as reported in S3.7). Parents gave written informed consent before participation, and infants received a small gift.

Stimuli and Procedure

Infants were presented with the same videos as adults (see Experiment 1). Infants saw a total of up to 40 videos per condition, presented in blocks of 20 videos of the same condition, in counterbalanced order across participants. On average, infants contributed 11.4 trials (± 6.1) in the Occluder condition and 12 trials (± 8.2) in the Tunnel condition. We monitored infants' attention online via camera. Each block started with the familiarization video that was played repeatedly until the infant had fully seen it once. Infants watched the videos on their parents' lap with a viewing distance of approximately 80 cm from the monitor.

Gaze coding

The experimental sessions were video-recorded for offline coding of infants' attention to the screen. Thirty percent of the videos were coded by a second independent rater. In case of disagreement, a third rater coded the data, and the rating of the majority was used.

EEG Recordings and SSVEP Analysis

We recorded a 28-channel EEG with the identical layout used for the adults. Because of strong baseline differences between conditions in the infant sample and an unbalanced number of trials between the social and no agent control conditions (BF_{10} = 1713.86, error estimation = 3.29×10^{-10}%), we used the signal-to-noise ratio (SNR) as a baseline-free measure of the SSVEP response that it is unbiased by different levels of noise. To extract the SNR, the trial averages were bandpass filtered using 10th order zero-phase shift IIR Butterworth filters with a halfwidth of 1 Hz on the frequency of interest (4 Hz) and on the neighboring frequencies (2 and 6 Hz). The SNR was calculated as the ratio between the signal envelope at 4 Hz divided by the average signal envelope at neighboring frequencies (2 and 6 Hz).

References

1. J. Perner, *Understanding the Representational Mind.* (The MIT Press, Cambridge, MA, US, 1991).

2. H. Rakoczy, Foundations of theory of mind and its development in early childhood. (2022).

3. B. Sodian, C. Thoermer, U. Metz, Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Dev. Sci.* **10**, 199–204 (2007).

4. Y. Luo, S. C. Johnson, Recognizing the role of perception in action at 6 months. *Dev. Sci.* **12**, 142–149 (2009).

5. S. Liu, T. D. Ullman, J. B. Tenenbaum, E. S. Spelke, Ten-month-old infants infer the value of goals from the costs of actions. *Science* **358**, 1038–1041 (2017).

6. U. Liszkowski, M. Carpenter, M. Tomasello, Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* **108**, 732–739 (2008).

7. J. Yott, D. Poulin-Dubois, Are Infants' Theory-of-Mind Abilities Well Integrated? Implicit Understanding of Intentions, Desires, and Beliefs. *J. Cogn. Dev.* **17**, 683–698 (2016).

8. F. Ting, Z. He, R. Baillargeon, Five-month-old infants attribute inferences based on general knowledge to agents. *J. Exp. Child Psychol.* **208**, 105126 (2021).

9. R. M. Scott, R. Baillargeon, Early False-Belief Understanding. *Trends Cogn. Sci.* **21**, 237–249 (2017).

10. D. Poulin-Dubois, H. Rakoczy, K. Burnside, C. Crivello, S. Dörrenberg, K. Edwards, H. Krist, L. Kulke, U. Liszkowski, J. Low, J. Perner, L. Powell, B. Priewasser, E. Rafetseder, T. Ruffman, Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cogn. Dev.* **48**, 302–315 (2018).

11. C. C. Hudson, A. L. Shamblaw, G. A. Wilson, M. M. Roes, M. A. Sabbagh, K. L. Harkness, Theory of Mind, Excessive Reassurance-Seeking, and Stress Generation in Depression: A Social-Cognitive-Interpersonal Integration. *J. Soc. Clin. Psychol.* **37**, 725–750 (2018).

12. J. Jara-Ettinger, P. Rubio-Fernandez, Quantitative mental state attributions in language understanding. *Sci. Adv.* **7**, eabj0970 (2021).

13. J. Jara-Ettinger, P. Rubio-Fernandez, The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychol. Rev.* **129**, 1394–1413 (2022).

14. R. P. R. D. Van Der Wel, N. Sebanz, G. Knoblich, Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition* **130**, 128–133 (2014).

15. D. Kampis, Á. M. Kovács, Seeing the World From Others' Perspective: 14-Month-Olds Show Altercentric Modulation Effects by Others' Beliefs. *Open Mind* **5**, 189–207 (2022).

16. Á. M. Kovács, E. Téglás, A. D. Endress, The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science* **330**, 1830–1834 (2010).

17. E. Ward, G. Ganis, P. Bach, Spontaneous Vicarious Perception of the Content of Another's Visual Perspective. *Curr. Biol.* **29**, 874-880.e4 (2019).

18. D. Kampis, E. Parise, G. Csibra, Á. M. Kovács, Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proc. R. Soc. B Biol. Sci.* **282**, 20151683 (2015).

19. V. Manea, D. Kampis, C. Grosse Wiesmann, B. Revencu, V. Southgate, An initial but receding altercentric bias in preverbal infants' memory. *Proc. R. Soc. B Biol. Sci.* **290**, 20230738 (2023).

20. C. Heyes, False belief in infancy: a fresh look. *Dev. Sci.* **17**, 647–659 (2014).

21. J. Phillips, D. C. Ong, A. D. R. Surtees, Y. Xin, S. Williams, R. Saxe, M. C. Frank, A Second Look at Automatic Theory of Mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychol. Sci.* **26**, 1353–1367 (2015).

22. V. Southgate, Are infants altercentric? The other and the self in early social cognition. *Psychol. Rev.* **127**, 505–523 (2020).

23. I. Santiesteban, C. Catmur, S. C. Hopkins, G. Bird, C. Heyes, Avatars and arrows: Implicit mentalizing or domain-general processing? *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 929–937 (2014).

24. F. Di Russo, S. Pitzalis, T. Aprile, G. Spitoni, F. Patria, A. Stella, D. Spinelli, S. A. Hillyard, Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. *Hum. Brain Mapp.* **28**, 323–334 (2007).

25. M. Köster, A. Brzozowska, A. Bánki, M. Tünte, E. K. Ward, S. Hoehl, Rhythmic visual stimulation as a window into early brain development: A systematic review. *Dev. Cogn. Neurosci.* **64**, 101315 (2023).

26. L. Narici, V. Pizzella, G. L. Romani, G. Torrioli, Evoked a- and -rhythm in humans: a neuromagnetic study.

27. I. C. Fiebelkorn, Y. B. Saalmann, S. Kastner, Rhythmic Sampling within and between Objects despite Sustained Attention at a Cued Location. *Curr. Biol.* **23**, 2553–2558 (2013).

28. M. Plöchl, I. Fiebelkorn, S. Kastner, J. Obleser, Attentional sampling of visual and auditory objects is captured by theta-modulated neural activity. *Eur. J. Neurosci.* **55**, 3067–3082 (2022).

29. A. N. Landau, H. M. Schreyer, S. van Pelt, P. Fries, Distributed Attention Is Implemented through Theta-Rhythmic Gamma Modulation. *Curr. Biol.* **25**, 2332–2337 (2015).

30. J. M. Kilner, Y. Paulignan, S. J. Blakemore, An Interference Effect of Observed Biological Movement on Action. *Curr. Biol.* **13**, 522–525 (2003).

31. D. Kampis, V. Southgate, Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends Cogn. Sci.* **24**, 945–959 (2020).

32. D. Samson, I. A. Apperly, J. J. Braithwaite, B. J. Andrews, S. E. Bodley Scott, Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 1255–1266 (2010).

33. G. G. Cole, A. C. Millett, The closing of the theory of mind: A critique of perspective-taking. *Psychon. Bull. Rev.* **26**, 1787–1802 (2019).

34. C. Grosse Wiesmann, A. D. Friederici, T. Singer, N. Steinbeis, Implicit and explicit false belief development in preschool children. *Dev. Sci.* **20**, e12445 (2017).

35. C. Grosse Wiesmann, A. D. Friederici, T. Singer, N. Steinbeis, Two systems for thinking about others' thoughts in the developing brain. *Proc. Natl. Acad. Sci.* **117**, 6928–6935 (2020).

36. K. Rothmaler, P. Berger, C. Grosse Wiesmann, Timing matters: disentangling the neurocognitive sequence of mentalizing. *Trends Cogn. Sci.* **26**, 906–908 (2022).

37. K. M. Tan, A. L. Daitch, P. Pinheiro-Chagas, K. C. R. Fox, J. Parvizi, M. D. Lieberman, Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others. *Nat. Commun.* **13**, 1919 (2022).

38. P. Berger, A. D. Friederici, C. Grosse Wiesmann, Maturational Indices of the Cognitive Control Network Are Associated with Inhibitory Control in Early Childhood. *J. Neurosci.* **42**, 6258–6266 (2022).

39. C. Grosse Wiesmann, J. Schreiber, T. Singer, N. Steinbeis, A. D. Friederici, White matter maturation is associated with the emergence of Theory of Mind in early childhood. *Nat. Commun.* **8**, 14692 (2017).

40. C. C. Klein, P. Berger, T. Goucha, A. D. Friederici, C. Grosse Wiesmann, Children's syntax is supported by the maturation of BA44 at 4 years, but of the posterior STS at 3 years of age. *Cereb. Cortex* **33**, 5426–5435 (2023).

41. M. Tomasello, Why we cooperate. *Why We Coop.*, xviii, 206–xviii, 206 (2009).

42. G. Csibra, G. Gergely, Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 1149–1157 (2011).

43. E. Herrmann, J. Call, M. V. Hernàndez-Lloreda, B. Hare, M. Tomasello, Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science* **317**, 1360–1366 (2007).

44. M. Tomasello, The ultra-social animal. *Eur. J. Soc. Psychol.* **44**, 187–194 (2014).

45. N. Mani, M. S. Schreiner, J. Brase, K. Köhler, K. Strassen, D. Postin, T. Schultze, Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Dev. Sci.* **24**, e13097 (2021).

46. F. D. Schönbrodt, E.-J. Wagenmakers, M. Zehetleitner, M. Perugini, Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychol. Methods* **22**, 322–339 (2017).

47. Preregistration aspredicted adults https://aspredicted.org/DRQ_OQV. https://aspredicted.org/DRQ_OQV.

48. Preregistration aspredicted infants https://aspredicted.org/WDR_NSM. https://aspredicted.org/WDR_NSM.

49. F. D. Schönbrodt, E.-J. Wagenmakers, Bayes factor design analysis: Planning for compelling evidence. *Psychon. Bull. Rev.* **25**, 128–142 (2018).

50. D. H. Brainard, The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).

51. D. G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).

52. M. Köster, M. Langeloh, S. Hoehl, Visually Entrained Theta Oscillations Increase for Unexpected Events in the Infant Brain. *Psychol. Sci.* **30**, 1656–1663 (2019).

53. M. M. Müller, W. Teder, S. A. Hillyard, Magnetoencephalographic recording of steadystate visual evoked cortical activity. *Brain Topogr.* **9**, 163–168 (1997).

54. F. Di Russo, A. Martínez, S. A. Hillyard, Source Analysis of Event-related Cortical Activity during Visuo-spatial Attention. *Cereb. Cortex* **13**, 486–499 (2003).

55. JASP Team, JASP, version 0.18.3 (2024).

56. M. D. Lee, E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, ed. 1, 2014; https://www.cambridge.org/core/product/identifier/9781139087759/type/book).

Acknowledgments: We would like to thank the infant, parents, and adult participants who volunteered to participate in this study. Further, we would like to thank Ruth Faßbender, Marie Michael, Valentina Tast, Noa Wallraff, Laura Walz, and Zeynep Yenen for their support in data collection and video coding the data. Finally, we would like to thank Frederik Bartholin, Radek Cichy, Joshua Grant, Paul Grosse Wiesmann, Stefanie Höhl, Andreas Keil, Vadim Nikulin, and Jonathan Phillips for very helpful comments to previous versions of this manuscript.

Funding: This study was funded by the FAZIT foundation for AT and the German Research Foundation (DFG, project number GR 5421/1-2) to AT, KR und CGW

Author contributions:

Conceptualization: CGW, MK, KR, AT

Methodology: CGW, MK, KR, AT

Investigation: AT, KR

Analysis: AT, KR

Visualization: AT, KR, CGW

Funding acquisition: CGW, AT

Project administration: AT, CGW

Supervision: CGW, KR

Writing – original draft: CGW, AT

Writing – review & editing: CGW, AT, KR, MK

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data in accordance with data protection regulations will be made available before acceptance.