Observing others’ joint attention increases 9-month-old infants’ object encoding

Supplementary Information

Table of contents

Video examples, scripts for pre-processing the eye-tracking data, as well as pilot data and analysis scripts for the power analysis are openly assessable on the Open Science Framework (blinded links):

Experiment 1: https://osf.io/yfegm/?view_only=943e5edf231446258bdcb7d2107972fb
Experiment 2: https://osf.io/dp5cg/?view_only=46c4d2d2ef3d4377b907fca49a615150

S1 - Supplementary Information Video Stimuli

Additional Information Objects

As objects, we used pictures of abstract toys created for a study by Wahl, Michel, Pauen, & Hoehl (2013). From the overall pool of 160 pictures, we selected 32 objects that could be fitted into a square shape. All infants saw the same pairs of objects, whereby each individual object served equally often as novel and as familiar object in all conditions (see Figure S1). For the purpose of our study, we edited the original pictures as follows: First, we removed the background-layer and replaced it with a transparent background. Then, we fitted each object into a 260×260 pixels square shape format. The side lengths of the square shape were determined by the height of the smallest image. Objects that were not square-shaped initially, were stretched or compressed into the required format to ensure that all objects covered a similar area on the screen. As a consequence, some objects had a slightly different shape as in the study by Wahl and colleagues (e.g., left object in pair 1, Figure S1). We selectively adjusted the color of the objects to ensure that the luminance and saturation was similar between objects. For one object we changed the color entirely from black to blue, since this was the only black object in the selection of object, presumably making it less salient compared to the other, more colorful objects (right object in pairing 9, Figure S1). All editing was done by using Adobe Photoshop.

Additional Information Video Content

This section provides detailed information regarding our considerations during stimulus development. We describe all deviations from the stimuli used previous studies, explain our decision to deviate from these stimuli, and elaborate on why we did not expect the deviations to have an impact on infants' object encoding.

The first deviation from previously used stimuli was that we only showed one object in the encoding phase (the familiar object). This means, that infants did not see the novel object until it appeared in the preferential-looking phase. In other screen-based studies, the novel object had been visible in the encoding phase already, but the actor did not pay attention to it (e.g., Okumura, Kanakogi, Kanda, Ishiguro, & Itakura, 2013; Okumura, Kanakogi, Kobayashi, & Itakura, 2020; Theuring, Gredebäck, & Hauf, 2007). Since most previous studies have looked at the relation between infants' gaze following and subsequent object encoding, the presence of two objects in the encoding phase was particularly required to allow calculating a gaze-following difference score. With regard to our specific research question, our primary goal was to create an arrangement between actors and object that allowed both actors to look away from the object without looking at any other object. Moreover, we wanted to create a scenario in which both actors played the same role in the interaction and were provided with the same amount of visual information. This excluded the possibility of providing one of the actors with two toys (one on both sides), as opposed to the other actor only having visual "access" to only one toy (as, e.g., in Meng, Uto, & Hashiya, 2017). During the stimulus planning phase, we considered that the presence of only one object in the encoding phase may raise the concern that infants' novelty preference could be at ceiling when seeing the novel object in the subsequent preferential-looking phase. However, since previous object-processing studies with real interactive settings had been done with only one object in the encoding phase as well, we assumed that the paradigm should principally work with the novel object first introduced in the preferential-looking phase (Begus, Southgate, & Gliga, 2015; Cleveland, Schug, & Striano, 2007; Cleveland & Striano, 2007; Ishikawa, Yoshimura, Sato, & Itakura, 2019).

Another difference from previous studies was the actor-object positioning during the encoding phase. Due to the presence of two (instead of one) actors, the actors in our videos were not positioned in the center, but rather within the left and right third of the screen. Correspondingly, the object was not positioned in the left or the right area of the screen, but at the bottom center instead. In contrast to previous studies, where the objects were positioned in left-right arrangement in both phases (encoding and preferential looking), the left-right positioning of the objects in the preferential-looking phase represented a new visual arrangement compared to the central placement in the encoding phase. However, since previous studies had shown that object encoding depends on object identity rather than object location (Okumura, Kobayashi, & Itakura, 2016), we did not consider the novel arrangement in the preferential-looking phase to have an effect on infants' processing.

When planning the manipulation of third-party ostensive cues, we aimed to create a context that was (according to previous findings) rich enough to elicit object encoding, but at the same time lean enough to trace infants' object encoding back to purely *third-party* ostensive cues. A recent study by Okumura and collegues (2020) suggests that 9-month-old infants may need infant directed speech in addition to eye contact to allow processing of an object (but see, e.g., Cleveland & Striano, 2007). Despite this finding, we did not include a corresponding initial greeting phase between the two actors, since we were concerned that the pure sound of the word "hello" could diminish the purely third-party context by giving infants the feeling of being addressed themselves, and thereby increasing their responsiveness (see also Senju & Csibra, 2008). To provide another ostensive cue in addition to third-party eye contact, we included the turning of the actors' entire body (toward or away from one another) as attention-grabbing social motion prior to third-party eye contact.

Related to the previous point, another deviation from previous videos was that the actors were shown in back view in the initial non-social sequence, rather than showing them facing forward with lowered gaze (Okumura et al., 2013; Okumura, Kanakogi, Kobayashi, & Itakura, 2017; Okumura et al., 2020; Theuring et al., 2007). We decided on this initial position as a clear demonstration of third-party context. We would argue that the back-view in our study has an even stronger non-social meaning compared to the previously used lowered-gaze-sequence, which is why we did not consider this a relevant deviation with regard to infants' processing performance.

Even though the overall duration of eye contact was consistent with previously used videos (2 seconds), we split this sequence in two one-second phases: one face-to-face (or back-to-back) phase before the actors look toward (or away from) the object, and one corresponding phase afterwards. Other screen-based studies have only used one eye contact sequence in the beginning. We decided for this twofold eye contact sequence for two reasons: First, we aimed to generally increase the interactive dynamic between the two actors, and second, we wanted to highlight the relation between the actors in the end of the video. We did not expect the presence of two eye contact sequences to have an impact on infants' object encoding compared to other studies, because the minimum requirement for communicative cueing remained fulfilled (namely eye contact *before* the gaze shift toward the object). To ensure that infants had payed attention to this minimum requirement, we only included trials during which infants had looked at the *first* eye contact sequence and the gazing sequence.

Additional Information Video Creation

For both experiments, the actors were filmed individually in front of a green screen. To ensure consistent timing of actions between actors and trials, metronome clicks were played at 120 bpm while filming. If necessary, we corrected the action timing of each actor post hoc and frame-by-frame in Adobe Premiere. Filming the actors in front of a green screen ensured flexible and accurate positioning of the dyad partners. Moreover, it allowed us to control for color and luminance differences between and within videos. We used Adobe Premiere Pro for cutting and editing the videos. Adobe Premiere's Ultra Key tool was used to isolate the actors from the background and replace it with an even colored, grey background layer which was identical over all videos. We positioned the actors in such a way that their overall motions were centered around the same vertical axis. In Figure S2 (Experiment 1) and Figure S3 (Experiment 2) we illustrate the areas of interest and maximum areas that the actors' movements covered over all conditions and trials.

S2 – Supplementary Analyses

We ran some analyses in addition to the analyses described in the main document to better understand the impact of infants' overt attention on their encoding performance. All following analyses served exploratory purposes and were not planned in the pre-registration. All face AOIs were defined 1° visual angle larger than the maximum areas covering all possible head movements from all actors in Experiment 1 and 2, to ensure comparability between the two experiments. The resulting AOIs covered an area of 14.4° × 10.9° (see Figure S2 and S3).

Experiment 1

Overall attention during encoding. First, we investigated condition differences in infants' overall attention to the encoding videos. For this purpose, we conducted a GLMM for infants' total looking time to the screen during the encoding phase (including fixation data over the entire trial sequence). We included the same fixed and random effects in the model as in our main model for infants' novelty preference score. We found that overall attention to the stimuli did not vary statistically across condition. Neither the interaction between third-party eye contact (eye contact, no eye contact) and others' looking at the object (looking at object, not looking at object) revealed a significant effect ($\chi^2(1) = 0.06$, $p = .80$, estimate = 96.8, $SE = 399.9$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 0.29$, $p = .59$, estimate = −104.1, $SE = 192.0$; looking at the object: $\chi^2(1) = 2.08$, $p = .15$, estimate = −284.5, $SE = 194.7$). However, infants' looking time to the familiarization videos decreased over trials (main effect of trial: $\chi^2(1) = 32.25$, $p < .001$, estimate = −1449.8, $SE = 194.3$). We found the same effect when repeating our model for infants' total looking times in the preferential looking phase (main effect of trial: $\chi^2(1) = 41.24$, $p < .001$, estimate = −1291.6, $SE = 140.8$), indicating a general decrease in visual attention throughout the experiment.

Looking times to the object and faces during encoding. To complement our analyses regarding the necessity of direct attention for object encoding, we assessed condition differences in infants' attention to the object in the encoding phase. For this purpose, we repeated the model of our main analysis for infants' fixation duration within the object AOI. We included the same fixed and random effects as in our main model. We found a main effect of others' looking at the object in that infants looked longer to the object when the actors did *not* look to the object ($\chi^2(1) = 10.05$, $p = .002$, estimate = 315.10, $SE = 93.46$). In addition, infants' attention to the object decreased over trials (main effect of trial: $\chi^2(1) = 4.22$, $p = .04$, estimate = −215.03, $SE = 101.31$). One possible explanation for the main effect of others' looking at the object is that the faces of the actors were systematically less visible in these conditions, meaning that they carried less information and social salience. As a possible consequence, the object may have received relatively more attention compared to the two conditions in which the actor's faces were visible. To explore this possibility further we repeated our analysis over the 5-second gazing phase only (i.e., the "still" sequence during which the actors looked away from or toward the object), revealing the same main effect of others' looking at the object ($\chi^2(1) = 8.84$, $p = .003$, estimate = 244.55, $SE = 81.36$). In addition, we compared infants' attention to the faces during the gazing phase. For this purpose, we calculated the sum of fixation durations within the two face AOIs for the corresponding sequence. We found a reversed main effect of others' looking at the object, in that infants looked longer to the faces in conditions during which the actors looked to the object ($\chi^2(1) = 17.68$, $p < .001$, estimate = −705.52, $SE = 149.53$). The pattern was the same when including fixations over the entire duration of the encoding phase ($\chi^2(1) = 10.39$, $p = .001$, estimate = −605.70, $SE = 176.90$). Table S3a provides a summary of the descriptive statistics for looking times during the encoding phase in all four conditions.

Taken together, our findings regarding infants' looking times suggest the following pattern: When the faces of the actors were visible, infants looked longer at the socially salient faces and shorter to the object. When the actors turned away from the object, their faces were less visible, causing longer looking times to the object and shorter looking times to the faces.

Gaze shifts to the object and between the faces during encoding. In addition to looking times, we explored infants' scanning pattern while they watched the videos in the encoding phase. First, we examined potential condition differences in the number of gaze shifts between the two faces of the actors. For this purpose, we determined the number of looks within each of the face AOIs. A "look" was defined as the interval between the first fixation on the active AOI and the end of the last fixation within the same active AOI when there were no fixations outside the AOI (Tobii Studio User Manual, Version 3.2, see also Meng et al., 2017). According to this definition one look could entail a group of multiple fixations. As a next step, we determined for each look whether the latest previous fixation (i.e., the last fixation immediately before the first fixation within the look) had hit the respective other face AOI. If this was the case, we counted this gaze event as a gaze-shift from one to the other AOI. If not (i.e., if the previous fixation before a look had been somewhere else on the screen), the corresponding look was labeled as irrelevant and discarded from the analysis. To examine condition differences, we conducted a GLMM for infants' total number of gaze shifts between the two faces, including all fixation data over the interaction phases (i.e., during the still face-to-face and back-to-back sequences). We included the same fixed and random effects as in our main model for infants' novelty preference score. Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect ($\chi^2(1) = 0.36$, $p = .55$, estimate = 0.09, $SE = .11$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 0.29$, $p = .59$, estimate = 0.05, $SE = 0.09$; looking at the object: $\chi^2(1) = 1.67$, $p = .20$, estimate = 0.10, $SE = 0.08$). The pattern remained the same when including fixation data over the entire duration of the video. This is in contrast to previous studies showing an increased number of gaze shifts between facing dyads as compared to people standing back-to-back (Augusti, Melinder, & Gredebäck, 2010; Meng et al., 2017). One possible explanation for the equal number of gaze shifts between the faces in the back-to-back conditions of our study is that infants were seeking information that could explain why the actors had turned away from one another in the first place.

In addition to gaze shifts between the two actors, we explored infants' object looks further. Specifically, we aimed to examine whether the origin of a specific look had an impact on infants' encoding performance, such that a socially caused referential look might be more relevant and therefore increase infants' processing compared to a look without any social origin. To test this assumption, we first calculated all looks within the object AOI, proceeding as described for the face-to-face gaze-shift analysis above. Then, we decided for each look whether the latest previous fixation had hit one of the two face AOIs. If this was the case, this look was labeled as socially-caused referential look. If the previous latest fixation did not hit any of the two face AOIs, the corresponding look was labeled as not socially-caused and discarded from the analysis. A trial was discarded from the analysis if no gaze shift had been performed toward the object at all. To examine condition differences, we conducted a GLMM for the total number of socially-caused object looks, including all fixation data of the gazing phase (i.e., the still sequence during which the actors looked at the object or away from the object). Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect ($\chi^2(1) = 1.51$, $p = .22$, estimate = 0.22, $SE = 0.18$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 2.33$, $p = .13$, estimate = 0.14, $SE = 0.09$; looking at the object: $\chi^2(1) = 0.05$, $p = .82$, estimate = −0.02, $SE = 0.09$). Moreover, in an additional analysis including fixation data over the entire video sequence, we did not find any significant correlation between the number of socially-caused object looks in the encoding phase and infants novelty preference score in the subsequent preferential-looking phase ($r(441) = -.01$, $p = .89$, $R^2 = .0001$). These results speak against the assumption that the observed triadic joint attention situation had increased infants' own attention to the object and thereby deepened their encoding of the object. Table S3b provides a summary of the descriptive statistics for gaze shifts during the encoding phase in all four conditions.

Taken together, we could not find any indication that infants' overt scanning pattern in the encoding phase of Experiment 1 had caused their increased processing in the third-party joint attentional condition. We did not find any evidence for increased gaze shifts between the two actors while they faced each other, nor did we find that the actors' gazing to the object had a direct impact on infants' own attention to the object.

Experiment 2

Overall attention during encoding. In contrast to Experiment 1, we found that infants' overall attention to the stimuli varied across conditions. Trials during which the actor looked to the object captured more global attention compared to trials during which the actor looked away from the object (main effect of others' looking at the object: $\chi^2(1) = 7.29$, $p = .007$, estimate = −545.6, $SE = 190.3$). One conceivable explanation for the difference between the experiments is that the videos in which the actors looked away from the object may have been less interesting in Experiment 2 compared to Experiment 1. Even though the visual appearance of the separate actors was the same in both Experiments, the presence of two actors looking to the side may have been more interesting compared to one actor performing the identical movement. As in Experiment 1, we found a continuous decrease in infants' looking time throughout the experiment—both in the encoding phase (main effect of trial: $\chi^2(1) = 21.50$, $p < .001$, estimate = −1224.6, $SE = 221.2$), as well as in the preferential-looking phase (main effect of trial: $\chi^2(1) = 29.31$, $p < .001$, estimate = −1102.1, $SE = 159.1$).

Looking times to the object and face during encoding. Neither the interaction between eye contact and others' looking at the object ($\chi^2(1) = 2.39$, $p = .12$, estimate = −253.85, $SE = 163.66$), nor the main effects of these two factors had a significant effect on infants' looking time to the object (eye contact: $\chi^2(1) = 1.01$, $p = .31$, estimate = −83.20, $SE = 82.24$; looking at the object: $\chi^2(1) = .46$, $p = .50$, estimate = 56.10, $SE = 82.21$). A main effect of trial indicated that infants' attention to the object decreased over trials ($\chi^2(1) = 19.04$, $p < .001$, estimate = −286.44, $SE = 56.15$). We did not find any systematic pattern regarding infants' looking duration to the faces either. Neither the interaction between eye contact and others' looking at the object ($\chi^2(1) = .09$, $p = .77$, estimate = 157.88, $SE = 539.75$), nor the main effects of these two factors revealed a significant effect on infants' looking time to the actor's face (eye contact: $\chi^2(1) = .04$, $p = .85$, estimate = 50.92, $SE = 269.81$; looking at the object: $\chi^2(1) = 1.24$, $p = .27$, estimate = −300.64, $SE = 269.81$). Overall, infants' attention to the faces decreased over trials ($\chi^2(1) = 6.93$, $p = .009$, estimate = −355.74, $SE = 135.07$). To ensure consistency between Experiment 1 and 2, we repeated our analyses over the 5-second gazing phase only. Infants' attention patterns remained the same for both the object as well as the actor's face. Table S4a provides a summary of the descriptive statistics for looking times during the encoding phase in all four conditions.

Gaze shifts to the object during encoding. As in Experiment 1, we examined the origin of infants' looks to the object further. We ran the same analyses as described in Experiment 1, and conducted the same pre-processing steps to extract socially caused looks at the object. Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect on the number of socially-caused looks in the gazing phase ($\chi^2(1) = 0.06$, $p = .81$, estimate = 0.04, $SE = 0.16$), nor did the main effects of the two factors (eye contact: $\chi^2(1) = 0.001$, $p = .99$, estimate = −0.001, $SE = 0.09$; looking at the object: $\chi^2(1) = 2.48$, $p = .12$, estimate = −0.12, $SE = 0.08$). Moreover, including fixation data from the entire video sequence, we did not find any significant correlation between the number of socially-caused object looks and infants' novelty preference score in the subsequent preferential-looking phase ($r(436) = −.01$, $p = .82$, $R^2 = .0001$). Given the previous literature on gaze following, one could have assumed that infants perform more socially-caused looks while the actor looked at the object. However, since we only presented one object in the encoding phase (rather than two objects as in previous gaze following studies), we could not directly calculate the difference score that has been previously used as standard measure of gaze following. One possible explanation for our finding is that there was no other visual stimulation on screen (such as a second object in previous gaze following studies), causing a continuous back-and forth looking between head and object regardless of condition. Table S4b provides a summary of the descriptive statistics for gaze shifts during the encoding phase in all four conditions.

Taken together, we did not find any indication from infants' overt looking behavior during the encoding phase of Experiment 2 (including looking time and gaze-shift measures) that may account for their increased object encoding performance in the critical joint attention condition.

Merged Analyses Experiment 1 and 2

Overall looking times during encoding. Infants' overall attention to the stimuli in the encoding phase did not differ between the two Experiments (Experiment 1: *M* = 6413.436, *SD* = 3001.84; Experiment 2: *M* = 6684.953; *SD* = 3146.908; χ^2 (1) = .46, *p* = .50, estimate = 277.5, *SE* = 410.7). This indicates that videos showing one person were equally interesting compared to videos showing two persons from a third-party perspective.

Relative attention to faces over objects during encoding. In Experiment 1 (Third-party), infants' proportional looking time to the faces over the object was significantly higher (*M* = .79, *SD* = .23) compared to Experiment 2 (*M* = .42, *SD* = .41; χ^2 (1) = 64.05, *p* < .001, estimate = −0.36, *SE* = 0.03). To calculate this proportion score, we divided infants cumulated fixation duration in the face AOIs by the sum of their cumulated fixation duration in the face AOIs and the object AOI. We included fixations from the total encoding video duration. The difference between the two experiments is not surprising since twice as many actors (and faces) were visible in in Experiment 1 compared to Experiment 2.

Overall, we did not find any systematic variation in infants' overt attention patterns that would suggest that infants' superior processing in the first- or third-party "eye contact – looking at object" condition depended on attention differences during encoding. This provides further support for the assumption that infants' object encoding had been driven by covert attentional processes.

Supplementary Information Fixation Filter

To define fixations, we used the Tobii Velocity-Threshold Identification (I-VT) fixation filter with default parameter values, that is: a velocity and distance threshold of 30° per second, no noise reduction, a maximum time between fixations of 75 ms, a maximum angle between fixations of 0.5°, a minimum fixation duration of 60 ms, and an interpolated of missing data for data segments below 75 ms. More details on the I-VT Filter can be found here: https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/determining-the-tobii-pro-i-vt-fixation-filters-default-values.pdf

S3 - Pilot Study

We conducted a pilot study to ensure that the video stimuli, the timing of the procedure, and the overall duration of the experiment were adequate for infants in the required age range. In addition, we used the pilot data to run a simulation-based a priori power analysis to determine whether our planned sample size was sufficient to detect the expected effect size. Piloting took place in March 2020 under the same conditions as the final study took place. We tested a version with 24 trials before deciding on the 16-trial version. Piloting was finished as soon as we had determined an age at which infants remained attentive throughout the experiment, while being old enough to ensure that they were sensitive to third party-interactions based on previous findings.

Participants

Overall, $N = 21$ infants between 9 months, 8 days and 13 months, 22 days participated in the pilot study ($M = 346.7$ days, $SD = 51.2$ days). The participants were recruited from the same data base as the participants for the final sample. We started piloting with versions of Experiment 1 ($n = 18$), since this experiment represented the main focus of this study. As soon as we had addressed all procedural concerns and decided on a concrete participant age range, we finished piloting for Experiment 1 and piloted three more infants in a corresponding version of Experiment 2. This was done to rule out that infants would be less attentive when only one person was visible on screen, and to check whether our piloting decisions based on Experiment 1 could also account for Experiment 2. For the a priori Power analysis, we only included infants who had participated in Experiment 1, and who provided at least one valid trial per condition after being filtered according to our criteria described in Experiment 1 in the main manuscript. This applied to *n* = 10 infants between 9 months, 8 days and 13 months, 22 days (*M* = 334.0 days, *SD* = 52.38 days).

A Priori Power Analysis

Since we piloted different versions of Experiment 1, some pilot participants were presented with 24 trials instead of 16 trials. To make best use of all data while ensuring consistency with our finally aimed data structure, we included the first 16 trials of these children in the power analysis, if they provided the sufficient number of one valid trial per condition. Due to the counterbalancing of conditions within trial-blocks, the first 16 trials in the 24-trial-version of the experiment included 4 trials of each condition, consistent with the final Experiment version.

We ran a simulation-based power analysis with the R package “simr” (Version 1.0.5, Green & MacLeod, 2019). The power analysis script and the pilot data are available online (blinded link: https://osf.io/yfegm/?view_only=943e5edf231446258bdcb7d2107972fb). We followed the following steps described by Green & MacLeod (2016). First, we fitted the main model of Experiment 1 (see analysis section of Experiment 1 in the main manuscript for details). The model estimates were calculated based on our pilot data. Second, we specified the effect size. As effect size, we calculated R-squared as the difference in the model fit between a full model including all fixed and random effects, and a null model including only the control variables and random effects without the fixed effects. R-squared was calculated with the R package “MuMIn” (Barton, 2019). The model comparisons revealed a marginal effect size of $R^2 = .077$ (conditional $R^2 = .089$). To detect an effect of this magnitude with a sample size of $N = 32$, the power analysis based on 1000 simulations indicated a power of 100% CI [99.63, 100.0]. We did not run a separate power analysis for the pilot version of Experiment 2, since the number of data points was too low to calculate a valid estimation of effect sizes. However, due to the closely matched study design, we expected a similar power in both experiments.

References

Augusti, E.-M., Melinder, A., & Gredebäck, G. (2010). Look who's talking: Pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00161

Barton, K. (2019). *Mu-MIn: Multi-model inference. R Package*. Version 1.40.0. https://cran.r-project.org/web/packages/MuMIn

Begus, K., Southgate, V., & Gliga, T. (2015). Neural mechanisms of infant learning: Differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biology Letters*, *11*, 20150041. https://doi.org/10.1098/rsbl.2015.0041

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5- and 7-month-old infants. *Infant and Child Development*, *16*, 295–306. https://doi.org/10.1002/icd.508

Cleveland, A., & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior and Development*, *30*, 499–504. https://doi.org/10.1016/j.infbeh.2006.10.009

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498. https://doi.org/10.1111/2041-210X.12504

Ishikawa, M., Yoshimura, M., Sato, H., & Itakura, S. (2019). Effects of attentional behaviours on infant visual preferences and object choice. *Cognitive Processing*, *20*, 317–324. https://doi.org/10.1007/s10339-019-00918-x

Meng, X., Uto, Y., & Hashiya, K. (2017). Observing Third-Party Attentional Relationships Affects Infants' Gaze Following: An Eye-Tracking Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.02065

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). The power of human gaze on infant learning. *Cognition*, *128*, 127–133. https://doi.org/10.1016/j.cognition.2013.03.011

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2017). Individual differences in object-processing explain the relationship between early gaze-following and later language development. *Cognition*, *166*, 418–424. https://doi.org/10.1016/j.cognition.2017.06.005

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. *Cognition*, *195*, 104082. https://doi.org/10.1016/j.cognition.2019.104082

Okumura, Y., Kobayashi, T., & Itakura, S. (2016). Eye Contact Affects Object Representation in 9-Month-Old Infants. *PLOS ONE*, *11*, e0165145. https://doi.org/10.1371/journal.pone.0165145

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*, 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Theuring, C., Gredebäck, G., & Hauf, P. (2007). Object processing during a joint gaze following task. *European Journal of Developmental Psychology*, *4*, 65–79. https://doi.org/10.1080/17405620601051246

Wahl, S., Michel, C., Pauen, S., & Hoehl, S. (2013). Head and eye movements affect object processing in 4-month-old infants more than an artificial orientation cue. *British Journal of Developmental Psychology*, *31*, 212–230. https://doi.org/10.1111/bjdp.12001

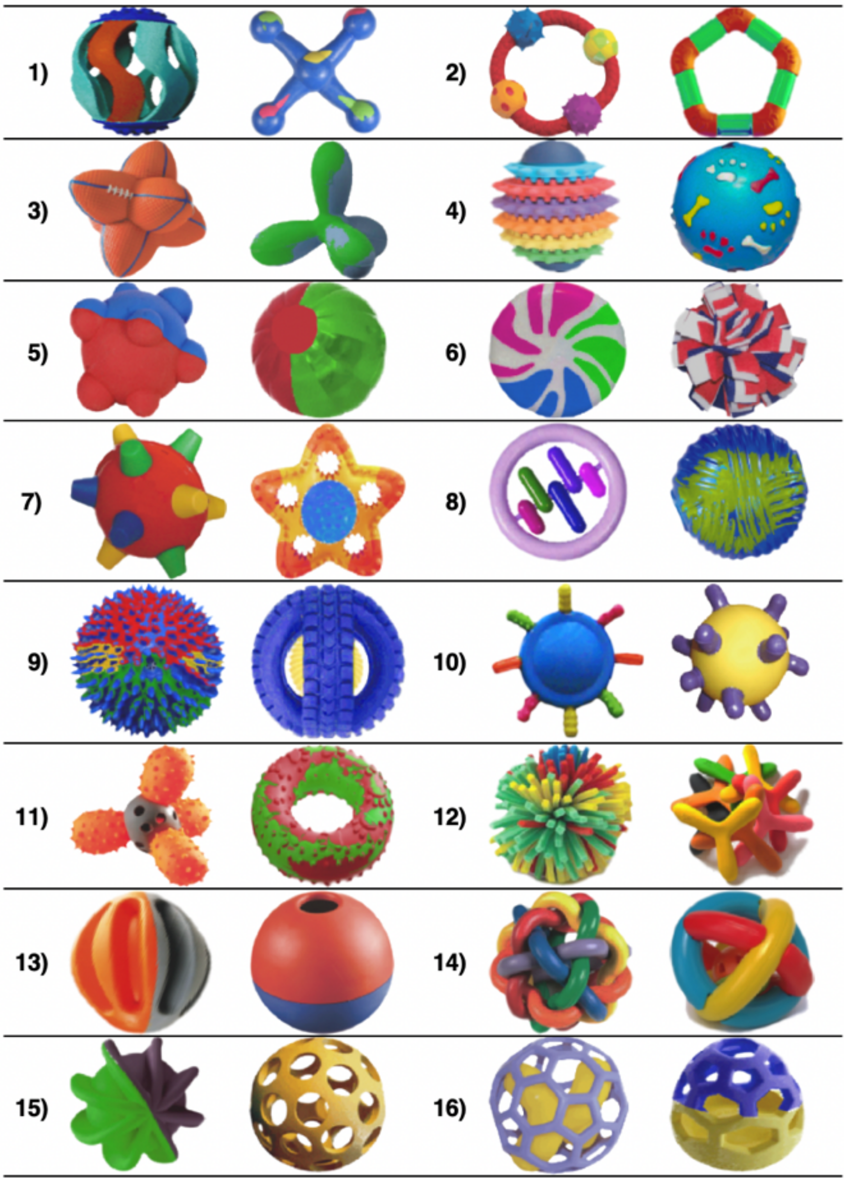

Figure S1. Overview of the 16 object-pairs that were used in the object-processing task in Experiment 1 and Experiment 2. The objects had been created for the study by Wahl et al. (2013). Each object served equally often as familiar object and as novel object in all four conditions.

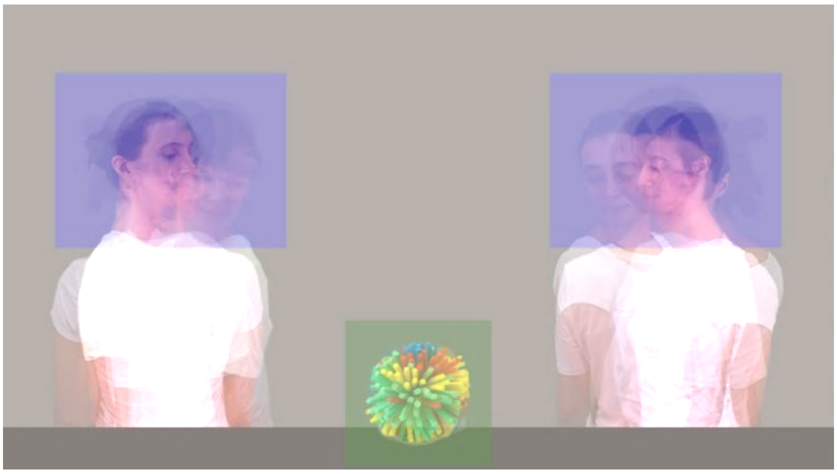

Figure S2. Areas of interest (AOIs) during the encoding phase of Experiment 1. All videos were presented in full-screen view (1920×1080 pixels). Green area = Object AOI (340×340 pixels), defined 1° visual angle larger than the maximum dimensions of the object. Blue areas = Face AOIs covering all possible head movements (570×430 pixels), defined 1° visual angle larger than the areas covering all possible head movements from all actors in Experiment 1 and 2.

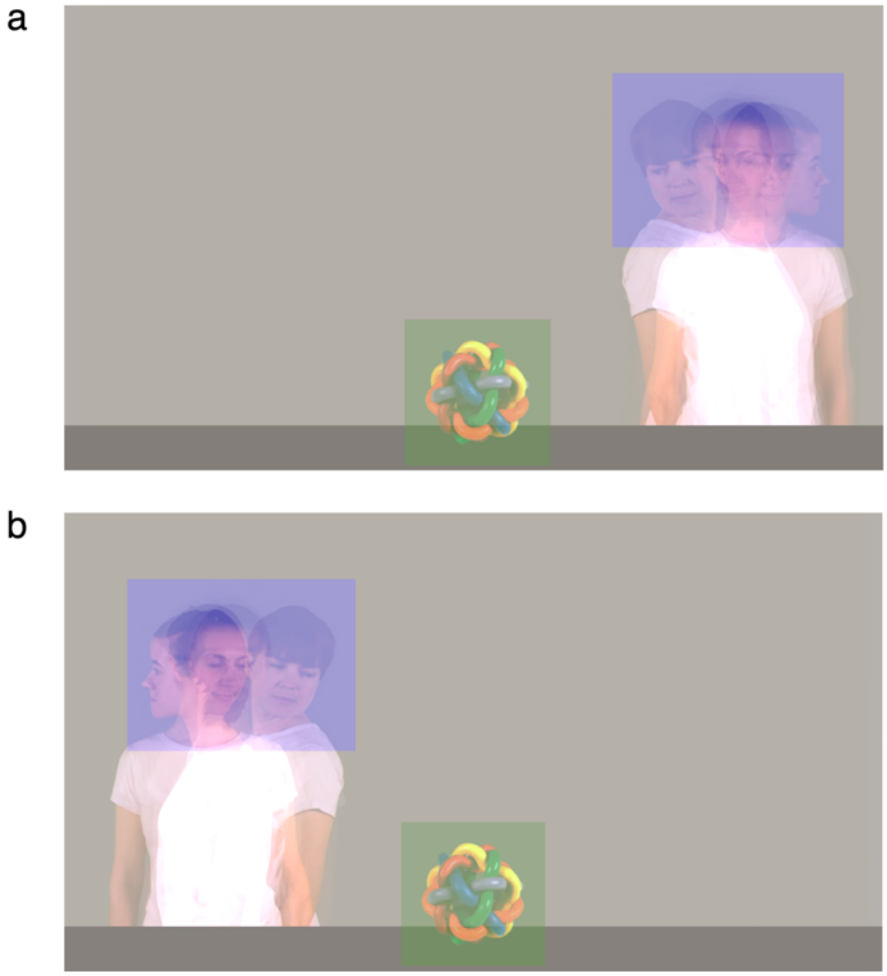

Figure S3. Areas of interest (AOIs) during the encoding phase of Experiment 2, (a) during trials showing the actor the right side of the object, and (b) during trials showing the actor on the left side. All videos were presented in full-screen view (1920×1080 pixels). Green area = Object AOI (340×340 pixels), defined 1° visual angle larger than the maximum dimensions of the object. Blue area = Face AOIs covering all possible head movements (570×430 pixels), defined 1° visual angle larger than the areas covering all possible head movements from all actors in Experiment 1 and 2.

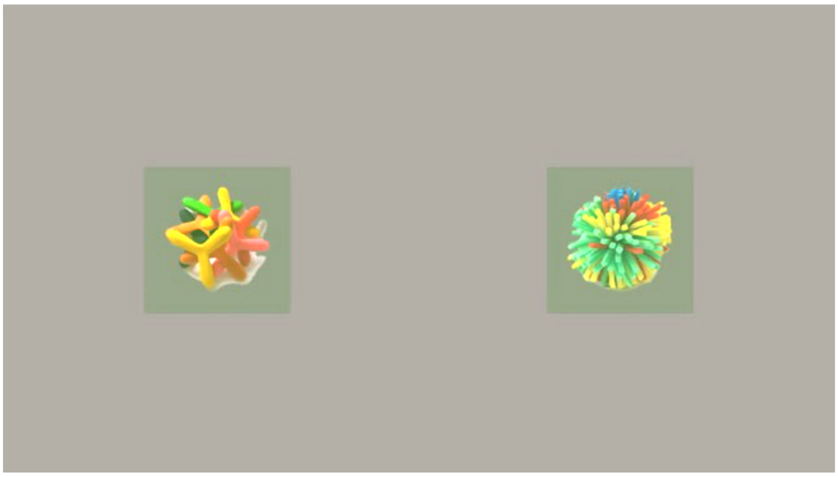

Figure S4. Areas of interest (AOIs) during the preferential-looking phase in Experiment 1 and Experiment 2. All videos were presented in full-screen view (1920×1080 pixels). Object AOIs (340×340 pixels) were defined 1° visual angle larger than the maximum dimensions of the objects.

Table S1

Valid trial statistics for the four conditions in Experiment 1

	Number of valid trials					
Condition	Total	Min	Max	M*(SD)*	Infant looking at object (fam.)	Infant not looking at object (fam.)
Third-Party Eye Contact Looking at object	111	2	4	3.47*(.67)*	73	38
No Third-Party Eye Contact Looking at Object	108	1	4	3.38*(.71)*	63	45
Third-Party Eye Contact No Looking at Object	110	2	4	3.44*(.76)*	82	28
No Third-Party Eye Contact No Looking at Object	114	2	4	3.56*(.62)*	82	32
TOTAL	443				300	143

Notes. Each individual infant could provide between 1 (min.) and 4 (max.) trials per condition. The maximum number of total trials over all infants was 128 per condition. The “infant looking at object (fam.)” column represents the number of valid trials during which infants had looked at the object at all over the total duration of the encoding phase (i.e., fixation duration > 0 ms within the object AOI). The “infant not looking at object (fam.)” column represents the number of valid trials during which infants had not looked at the object at all over the total duration of the encoding phase (i.e., fixation duration = 0 within the object AOI).

Table S2

Valid trial statistics for the four conditions in Experiment 2

	Number of valid trials					
Condition	Total	Min	Max	M*(SD)*	Infant looking at object (fam.)	Infant not looking at object (fam.)
First-Party Eye Contact Looking at Object	108	1	4	3.38*(.91)*	89	19
No First-Party Eye Contact Looking at Object	109	2	4	3.41*(.76)*	100	9
First-Party Eye Contact No Looking at Object	109	2	4	3.41*(.76)*	93	16
No First-Party Eye Contact No Looking at Object	112	1	4	3.50*(.84)*	102	10
TOTAL	438				384	54

Notes. Each individual infant could provide between 1 (min.) and 4 (max.) trials per condition. The maximum number of total trials over all infants was 128 per condition. The “infant looking at object (fam.)” column represents the number of valid trials during which infants had looked at the object at all over the total duration of the encoding phase (i.e., fixation duration > 0 ms within the object AOI). The “infant not looking at object (fam.)” column represents the number of valid trials during which infants had not looked at the object at all over the total duration of the encoding phase (i.e., fixation duration = 0 within the object AOI).

Table S3

Means (and standard deviations) for looking times and gaze-shifts during the encoding phase in all four conditions of Experiment 1

	Conditions Experiment 1 (Third-party)			
	Eye Contact/ Looking at object	No Eye Contact/ Looking at object	Eye Contact/ No Looking at object	No Eye Contact/ No Looking at Object
(a) Looking times				
LT Screen (Overall)	6458.62 *(1708.40)*	6598.85 *(1817.30)*	6215.88 *(1506.73)*	6380.40 *(1575.40)*
LT Object AOI (Overall)	1087.98 *(905.48)*	966.64 *(984.07)*	1335.97 *(1072.26)*	1399.38 *(1174.15)*
LT Object AOI (Gazing)	779.77 *(657.61)*	668.98 *(711.82)*	943.73 *(673.56)*	1022.40 *(751.51)*
LT Face AOIs (Overall)	4875.75 *(1894.56)*	5060.34 *(1791.05)*	4393.43 *(1728.33)*	4421.97 *(1661.28)*
LT Face AOIs (Gazing)	2951.35 *(1391.08)*	3171.61 *(1211.17)*	2390.10 *(1213.94)*	2415.97 *(1082.29)*
(b) Gaze shifts				
Gaze shifts between two face AOIs (Overall)	1.95 *(1.28)*	1.97 *(1.20)*	1.88 *(1.27)*	1.63 *(1.24)*
Gaze shifts between two face AOIs (Interaction)	0.91 *(0.68)*	0.91 *(0.48)*	1.05 (*0.72)*	0.96 *(0.69)*
Socially-Caused Looks at the object (Gazing)	0.96 *(0.54)*	1.0 *(0.65)*	1.09 *(0.56)*	0.90 *(0.55)*
Overall Looks at the Object (Gazing)	1.63 *(0.67)*	1.63 *(0.63)*	1.74 *(0.63)*	1.68 *(0.58)*

Notes. In the first column in parentheses, “Overall” refers to the entire duration of the encoding phase (max. duration = 11000 ms), “Gazing” refers to the phase in which the actors looked toward or away from the object (max. duration = 5000 ms), “Interaction” refers to the phases in which the actors looked at each other’s eyes or away from one another (max.

duration = 2000 ms). (a) Looking times (LT) represent the sum of fixation durations within the corresponding area of interest (AOI) in milliseconds (ms). (b) Gaze shifts represent the total numbers of gaze movements between the two face AOIs, as well as gaze shifts toward the object. “Socially-Caused Looks at the Object” include gaze shifts from the face AOI to the object AOI, “Overall Looks at the Object” include all gaze shifts toward the object (i.e., both socially-caused and not socially-caused gaze shifts).

Table S4

Means (and standard deviations) for looking times and gaze shifts during the encoding phase in all four conditions of Experiment 2

	Conditions Experiment 2 (First-party)			
	Eye Contact/ Looking at object	No Eye Contact/ Looking at object	Eye Contact/ No Looking at object	No Eye Contact/ No Looking at Object
(a) Looking times				
LT Screen (Overall)	6854.07 *(2304.36)*	7018.86 *(2114.37)*	6721.62 *(2097.61)*	6145.27 *(1968.59)*
LT Object AOI (Overall)	1232.92 *(836.77)*	1203.86 *(751.38)*	1174.88 *(749.56)*	1369.24 *(890.84)*
LT Object AOI (Gazing)	929.87 *(610.74)*	861.02 *(598.35)*	792.21 *(572.78)*	1013.16 *(560.83)*
LT Face AOIs (Overall)	2591.37 *(1018.47)*	2635.04 *(1184.88)*	2391.37 *(1191.77)*	2241.02 *(1263.93)*
LT Face AOIs (Gazing)	1537.59 *(697.44)*	1652.01 *(873.90)*	1407.35 *(888.28)*	1180.11 *(898.04)*
(b) Gaze shifts				
Socially-Caused Looks at the object (Gazing)	1.33 *(0.57)*	1.31 *(0.49)*	1.22 *(0.48)*	1.17 *(0.42)*
Overall Looks at the Object (Gazing)	1.69 *(0.48)*	1.58 *(0.51)*	1.60 *(0.56)*	1.70 *(0.40)*

Notes. In the first column in parentheses, “Overall” refers to gaze events over the entire duration of the encoding phase (max. duration = 11000 ms) and “Gazing” refers to the phase in which the actor looked toward or away from the object (max. duration = 5000 ms). (a) Looking times (LT) represent the sum of fixation durations within the corresponding area of interest (AOI) in milliseconds (ms). (b) Gaze shifts represent the total numbers of gaze movements toward the object. “Socially-Caused Looks at the Object” include gaze shifts from the face AOIs to the object AOI, “Overall Looks at the Object” include all gaze shifts toward the object (i.e., both socially-caused and not socially-caused gaze shifts).