Joint attention biases dogs' memory towards object identity

Andrea Sommese[1, 2], Maleen Thiele[3], and Christoph J. Völter[1, 3]

[1]Comparative Cognition, Messerli Research Institute, University of Veterinary Medicine of Vienna, Medical University of Vienna and University of Vienna, Vienna 1210, Austria

[2]Centre for Animal Nutrition and Welfare, Clinical Department for Farm Animals and Food System Science, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Vienna, Austria

[3]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany.

Corresponding authors: Andrea Sommese; Christoph J. Völter

University of Veterinary Medicine Vienna, Veterinaerplatz 1, 1210 Vienna, Austria

Email: andrea.sommese@vetmeduni.ac.at; christoph_voelter@eva.mpg.de

ORCID: AS: 0000-0001-8859-5018; CV: 0000-0002-8368-7201

Author Contributions: AS: Conceptualization; Methodology; Project administration; Validation; Writing - original draft; Writing - review & editing. MT: Conceptualization; Methodology; Resources; Writing - review & editing. CJV: Conceptualization; Data curation; Formal analysis; Methodology; Validation; Visualization; Writing - review & editing; Funding acquisition; Resources; Supervision; Writing - original draft; Writing - review & editing.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences, Psychological and Cognitive Sciences.

Keywords: ostensive communication, joint attention, object encoding, eye-tracking, social learning, dog cognition.

Abstract

Ostensive communication, characterized by direct eye contact and other attention-catching signals, shapes how humans from an early age encode and remember novel objects, specifically biasing their memory toward identity-relevant features. Dogs are sensitive to human ostensive cues and reliably follow gaze direction after communicative addressing. However, whether these signals also enhance and modulate object encoding as in humans remains unknown. In a preregistered eye-tracking study, we tested pet dogs (N = 35) in a violation-of-expectation paradigm. A human actor either directed eye contact toward the dog or looked away before looking at an object. After an occlusion event, dogs viewed one of three outcomes: the same object in the same location (no change), the same object in a different location (location change), or a different object in the same location (identity change). We measured the dogs' looking times during the outcome phase within predefined areas of interest around the object. In line with our predictions and consistent with earlier findings with human infants, dogs looked significantly longer at identity changes in the communicative than in the no-eye-contact condition. In contrast, looking times at location and no-change outcomes were unaffected by communicative context, indicating a selective enhancement of identity encoding. During the initial addressing phase, pupil dilation was greater in the eye-contact condition, suggesting increased arousal or engagement. These findings demonstrate that dogs, like human infants, exhibit a communication-induced memory bias, revealing a shared capacity for ostension-guided learning that may facilitate information transfer between humans and dogs.

Significance Statement

Human infants selectively encode object identity features when addressed through ostensive communication involving cues such as eye contact that signal communicative intent. This study reveals that dogs possess this important learning mechanism, too. Using eye-tracking methodology, we show that, following ostensive human communication, dogs exhibit enhanced memory for object identity, but not location. This selective memory bias mirrors findings with human infants. These results reveal the existence of a shared capacity for ostension-guided learning in humans and dogs, providing insights into the cognitive mechanisms that facilitate effective interspecies communication and social learning.

Main Text

Introduction

Human infants possess a remarkable capacity to acquire culturally shared knowledge through social learning mechanisms that emerge within their first year of life (1, 2). Among the

key mechanisms is joint attention, the shared and mutually recognized attention between two individuals and an object (1–5). Joint attention is often established through ostensive signals, such as eye contact, verbal addressing (6, 7) or physical proximity and touch (8–11). The role of joint attention extends beyond attentional coordination to enhancing memory formation and learning. From four months onwards, infants show enhanced object encoding when they observe an adult looking at an object after being addressed ostensively (12–14), with their sensitivity to eye contact increasing across infancy (live interactions: 14, 15; screen-based settings: 16–18).

As infants progress into the second half of their first year, the influence of joint attention and communicative context on their memory formation becomes even more evident: they show selective encoding of identity-relevant object features (19–21). This "communication-induced memory bias" aligns with the Natural Pedagogy framework, which assumes that communicative signals bias the learner toward object features that are relevant for the learning of kind-relevant, generalizable information (22–24). Initial evidence for this bias in preverbal infants came from a study by Yoon et al. (21). In a screen-based setup, 9-month-olds who encountered an object in a joint attention interaction involving communicative cueing by an adult (eye contact, infant-directed speech, pointing), later looked longer when the object's identity had changed than when its location had changed, suggesting selective encoding of internal, identity-relevant over external, transient object features. Although not replicated in one study (25), subsequent work has demonstrated similar patterns using only eye contact to establish the communicative context in live interactions (20) and in screen-based settings (19). The latter study also provided evidence that the bias can arise through third-party observation (19). A similar communication-induced memory bias has also been found in human adults (26, 27).

Despite extensive research on human social learning, the evolutionary origins and cross-species generality of ostension-guided learning remain largely unexplored (28). Dogs represent an ideal comparative model, given their unique evolutionary history with humans and their demonstrated sensitivity to human communicative signals (29–39). Canine cognition research has shown that dogs track human attentional states (40, 41), exhibit visual perspective-taking abilities (29–31, 35, 42) and even monitor human knowledge states (33, 34, 38). Ostensive cues such as verbal addressing and eye contact reliably enhance dogs' gaze following and bias their choices (32, 33, 36, 37, 39, 43, 44). However, what remains unknown is whether these cues shape not only where dogs attend but also what they encode about objects, specifically, whether eye contact induces a human-like bias to prioritize identity over location.

Here, we tested this question in a preregistered eye-tracking study using the violation-of-expectation paradigm and video stimuli by Thiele et al. (19). Pet dogs viewed videos in which a woman either looked toward the dog (eye-contact condition) or looked away (no-eye-contact condition) before and after looking at an object. Following a brief occlusion, we presented one of three outcomes: the same object reappeared in the same location on screen (no change), the

same object appeared in a different location (location change), or a completely novel object appeared in the same location (identity change). If dogs share humans' preparedness for ostension-guided learning, we predicted that they should selectively encode recognition-relevant object properties and, consequently, look longer to identity changes, particularly following eye contact.

Results

To examine whether ostensive signals influence dogs' object encoding, we analyzed dogs' looking patterns and pupil size changes in an eye-tracking experiment using a violation-of-expectation paradigm. Confirmatory and exploratory analyses revealed that dogs looked longer and faster at objects that changed identity, specifically following eye contact. These effects were not observed for location changes or unchanged objects. Additional analyses showed increased pupil dilation during eye contact, supporting heightened arousal or engagement.

Confirmatory analysis

Dwell time in the outcome phase

Dogs spent significantly more time looking at objects when their identity changed, but only after the human actor had made eye contact with them (Fig. 1A, B). The full model provided a significantly better fit to the data compared to the null model, which only included the random effects ($\chi2(7) = 30.80$, $p < 0.001$). As predicted, we found a significant interaction between condition and outcome ($\chi2(2) = 7.08$, $p = 0.029$; Table S1). The difference between the eye-contact and no-eye-contact conditions in the dwell times was significantly greater in the identity change than in the no-change condition ($t = -2.19$, $df = 217.53$, $p = 0.029$). No significant interaction was found for the location change vs no change contrast ($t = -0.11$, $df = 213.88$, $p = 0.909$). We found that the dogs looked significantly longer at the object in the eye-contact than no-eye-contact condition following an identity change ($t = -2.35$, $df = 215.19$, $p = 0.020$) but not following a location change ($t = 0.64$, $df = 211.93$, $p = 0.524$) or no change ($t = 0.76$, $df = 212.04$, $p = 0.448$; Fig. 1A; see ESM for mean dwell times). Additionally, the dogs looked significantly more often at the upper object ($t = 4.81$, $df = 33.88$, $p < 0.001$). The trial number had no significant effect ($t = 0.85$, $df = 30.96$, $p = 0.400$).

Exploratory analysis

Latency to look at the object in the outcome phase

Dogs not only looked longer at identity changes following eye contact but also oriented to these changes more rapidly. The full model provided a significantly better fit to the data compared to the null model, which only included the random effects ($\chi2(7) = 22.54$, $p < 0.001$). As before, we found a significant interaction between condition and outcome ($\chi2(2) = 8.87$, $p = 0.012$; Table

S2). The difference between the eye-contact and no-eye-contact conditions was significantly greater in the identity change than in the no-change condition ($t = -2.49$, $df = 179.44$, $p = 0.014$). No significant interaction was found for the location change vs no change contrast ($t = -0.53$, $df = 181.31$, $p = 0.600$). We found that the dogs looked significantly faster at the object in the eye-contact than no-eye-contact condition following an identity change ($t = 2.10$, $df = 127.65$, $p= 0.038$) but not following a location change ($t = -0.63$, $df = 122.47$ $p = 0.533$) or no change ($t = -1.31$, $df = 134.40$, $p = 0.194$; Fig. 1C). Additionally, the dogs looked significantly sooner at the upper object ($t = -3.95$, $df = 27.51$, $p < 0.001$). The trial number had no significant effect ($t = -1.76$, $df = 208.10$, $p = 0.080$).

Dwell times to actor and object in the action phase

To ensure that our results reflected differential encoding rather than biased attention during initial stimulus presentation, we analyzed the dogs' looking behavior during the action phase when the human actor looked at the object. We found no evidence for a significant difference in attention during the action phase between conditions, neither in the dogs' looking times to the actor ($\chi2(1) = 0.54$, $p = 0.461$) nor the object ($\chi2(1) = 0.65$, $p = 0.422$).

Pupil size changes in the action phase

We analyzed the preprocessed, baseline-corrected pupil size data at the beginning of the action phase, in which the actor turned around looking toward the dog's direction or to the side (Fig. 1D; for the raw pupil size data, see Fig. S1 A-B) using a generalized additive mixed model (GAMM01). The full model, including condition, fitted the data significantly better than the null model (Chi-square test of ML scores: $\chi2(3) = 9.51$, $p < 0.001$; GAMM01 had a lower AIC: ΔAIC 26.96). The pupil size varied significantly over time in both conditions (eye-contact: $F(12.67, 14.88) = 4.74$, $p < 0.001$; no eye-contact: $F(13.14, 15.32) = 3.42$, $p < 0.001$). While over the entire interest period (3000 – 7000 ms) there was no significant difference between the dogs' pupil size in the eye-contact and no eye-contact condition ($t = -1.10$, $p = 0.271$), the difference curve revealed that the pupil size was significantly larger in the eye-contact condition than the no-eye-contact condition in the time window between 4697 and 5505 ms (Fig. 1E; for the model estimates and partial and summed effects see Table S4 and Fig. S2).

Figure 1.

Discussion

Our findings provide initial evidence that ostensive signals, specifically eye contact, selectively influence dogs' encoding of object identity, mirroring communicative learning biases reported in human infants (19–21). The pet dogs in our study oriented more rapidly and looked

longer at identity changes, but only when the actor first established eye contact. This suggests that, as in infants, ostensive cues direct information processing on intrinsic features of objects rather than transient spatial information.

These findings show that dogs' memory and attention can be shaped not merely by associative processes or stimulus salience, but also by communicative context. This effect is closely aligned with observations in preverbal infants, who also selectively encode identity-relevant properties of objects when interactions are marked by ostensive signals and interpersonal jointness (19–21). Our results, therefore, extend the scope of the Natural Pedagogy theory (22, 23) beyond humans and hominids (28). The theory originally proposed that humans are specifically prepared to use ostensive signals from an early age to learn generalizable information from others. Our data indicate that at least one domesticated, nonhuman species shows a similar selectivity under ostensive engagement. These findings are consistent with the notion that domestication and enculturation with humans have tuned dogs to treat direct eye contact in combination with head turns toward an object as cues to relevant and generalizable content as well.

In the context of a pointing task, an earlier study (45) suggested that dogs preferentially encode an object's location rather than its identity-relevant features following ostensive addressing. To examine this, the experimenter, after pointing at one of the objects, switched the locations of two objects visible to the dogs. After the switch, dogs preferentially approached the initially cued location rather than the originally cued object in the condition involving eye contact and verbal addressing. In contrast, our findings suggest that eye contact can bias dogs' memory toward recognition-relevant object features. Two non-exclusive explanations could resolve this discrepancy with our findings. First, pointing may be interpreted by dogs as an imperative, spatial directive. If so, pointing might bias dogs toward locations or directions, whereas eye contact alone can bias encoding toward object identity. Second, several studies have shown that ostensive communication can induce search errors in dogs even when they have witnessed where an object has actually been hidden (32, 33, 37, 43, 44), which could inflate location-based choices. Our paradigm avoided pointing, hiding, and choice demands, and we still observed an identity bias selectively after eye contact. Thus, when isolating ostension and minimizing task demands that potentially emphasize the spatial dimension or response habits, dogs exhibit a human-like bias toward identity-relevant features.

Pupillometry offers a window into potential mechanisms. Eye contact elicited a transient increase in pupil size, consistent with a brief arousal or engagement response, potentially facilitating encoding of referential information into memory. Indeed, in humans, pupil dilation during encoding of new information is predictive of successful later memory retrieval (46, 47). Additionally, in human infants, eye contact can lead to increased arousal (measured using heart rates) and predict subsequent responsiveness to referential cues (48, 49).

While these results indicate a parallel between dogs and infants in their sensitivity to ostensive cues, dogs likely still differ from human infants in how flexibly and broadly they generalize information, and the degree to which they rely on familiarity or training (37). Throughout their domestication, dogs might have undergone selection for socio-cognitive abilities that allow them to interpret and adequately respond to human communicative signals (37, 50). However, more research is needed to elucidate the boundaries and mechanisms underlying this capacity, and how it compares to those of other species, both domesticated and non-domesticated.

Although our study benefits from the accuracy and detailed data offered by eye-tracking technology, some limitations merit consideration. A primary concern is the use of screen-based stimuli, which differ from dogs' real-life interactions and therefore raise questions about ecological validity. However, evidence from studies with infants shows that communication-induced memory biases emerge both in screen-based procedures (19, 21) and in live interactive setups (20), suggesting that the underlying mechanism is consistent across contexts and not an artefact of presentation format. Screen-based research paradigms also indicate that dogs actively engage with and interpret social and communicative signals on screens. For instance, Téglás et al. (36) found that dogs follow human gaze cues presented on a screen, particularly after being addressed ostensively. Karl et al. (51) found that dogs looked longer at pictures of their caregivers on a screen than at another familiar person, and their pupils dilated more when the depicted person exhibited an angry facial expression. Further screen-based research shows that dogs can learn to categorize pictures (52), anticipate the destination of moving objects on the screen (53), react to animacy cues (53), and show sensitivity to certain physical regularities (53, 54; for a review, see 38). Together, these findings support the validity of using screen-based paradigms to investigate cognitive processes in both humans and dogs. In our study, we further controlled for passive viewing by excluding trials where dogs were already looking at the object's location at the onset of the outcome phase, ensuring that recorded gaze shifts reflected an active response to the social cue rather than passive viewing. Moreover, our experimental design closely followed the paradigm developed by Thiele et al. (19) for infants, employing the same video stimuli and trial structures to enable direct cross-species comparison.

Future research would benefit from testing how dogs' ostension-guided memory encoding unfolds in more naturalistic settings, using live social partners and incorporating multimodal cues such as gesture, vocalizations, prosody, and body orientation. Understanding how these cues interact, whether they are additive, redundant, or context-dependent, could offer further insight into the mechanisms by which dogs infer relevance and prioritize information. For instance, pointing gestures or vocal intonation may shift dogs' expectations toward particular types of content (e.g., spatial vs. identity-relevant information), potentially modulating what they encode (45). Factors such as familiarity, emotional expression, or perceived relevance may affect

how ostensive signals are interpreted and how readily they facilitate learning. Comparative studies examining the impact of social versus non-social attention-getters, such as sudden movement or sound, could also help disentangle whether the observed effects are specific to communicative intent or reflect broader attentional processes. Finally, individual differences likely influence how dogs respond to ostensive cues. Factors such as breed, training history, and life experience may shape sensitivity to communicative signals. For instance, working dogs may be more attuned to intentional cues due to training and companion breeds might show heightened responsiveness to eye contact because of selection for social engagement.

In summary, our results demonstrate that human communicative signals (eye contact) not only capture canine attention but can also bias which features of a scene are encoded and remembered, similar to mechanisms of early social learning seen in humans. These findings have practical implications for developing more effective training programs for pet and working dogs based on communicative engagement and joint attention. They also contribute to theoretical discussions on Natural Pedagogy, suggesting that some mechanisms of ostension-guided learning may be shared at least with a domesticated species. By aligning our experimental design with studies in human infants, we provide a comparative framework to explore the evolutionary origins of social cognition and highlight the potential for dogs to serve as models for understanding cross-species learning mechanisms.

Material and methods

The experimental design, hypotheses, predictions, sample size, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/YZ6AP). All analysis scripts and data associated with the study are openly available in the repository cited in the Data, Materials, and Software Availability section. The procedure described in this study was discussed and approved by the Ethics and Animal Welfare Committee of the University of Veterinary Medicine Vienna, following the university's guidelines for good scientific practice (ETK-179/11/2023).

Subjects

Our target sample size (N = 36) was determined based on the results of an a priori power simulation reported by Thiele et al. (19), which used the same design and stimuli in a study with human infants. The simulation estimated effect sizes using raw data from Okumura et al. (20). We started testing 36 pet dogs (18 females) of various breeds and mixes (Table S5 provides demographic information). The mean age at the beginning of the experiment was 56.1 months (±4.21; range 11–103 months). Of these, 34 dogs completed all trials, while one additional dog became unavailable after the first session but was still included in the analyses (final sample: N = 35).

Study design

Following the pretest, each dog completed 12 test trials structured according to a 2 × 3 factorial design (using the design by 18). The two manipulated factors were: (1) the presence of eye contact between the actor and the dog during the action phase (eye contact vs. no eye contact) and (2) object change during the outcome phase (no change, identity change, location change). Each dog experienced two trials per condition, for a total of six conditions. Trials were distributed across four sessions, each consisting of three trials with distinct outcome conditions.

Stimuli

We used the video stimuli created for the study by Thiele et al. (18), differing only in our selection of object images, which were taken from the same object collection that was initially created for a study by Wahl et al. (14). Each video consisted of 3 phases: (1) an action phase featuring a 15-second video of an actor gazing at a novel object; (2) a delay phase in which the scene was occluded for 3 seconds; and (3) an outcome phase where a single object was shown for 15 seconds or until the dog looked away from the screen for two consecutive seconds, following a minimum of 100 ms of gaze at the object's area of interest (AoI; 350 x 350 px).

During the **action phase**, a curtain opened (1s) to reveal one of two adult women on either side of the screen (actor AoI: 550 × 760 px). The actor's position (left or right) was fully counterbalanced across dogs, eye-contact conditions, and outcome conditions. Each dog saw the actor three times on each side within each condition. For the first two test trials, the actor's position was counterbalanced so that half of the dogs saw her on the left first, and the other half on the right, with the opposite position in the second trial. The actor was initially shown in back view (1s). In the eye-contact condition, the actor first turned toward the dog (1s) and looked at the dog for (1s). Then, she shifted her gaze during 1s to a centrally visible object (positioned at the top or bottom of the screen, counterbalanced), looked at it (1s), and returned her gaze to the dog during 1s, looking again at the dog (1s). This alternating gaze shift between the dog and the object was repeated twice. In the no-eye-contact condition, timing and body movements were identical, but the actor was seen in a side-view, facing outward and never establishing eye contact with the dog. Instead, she alternated her gaze between the object and her original side-facing position. In both conditions, the actor maintained a neutral facial expression and did not speak. The action phase ended with the curtain closing (1s). During the **delay phase,** the screen remained occluded by the curtain (3s). Finally, in **the outcome phase**, the curtain reopened (1s) to reveal one of three outcomes: (1) the same object in the same screen location (no change), (2) the same object in a new location (location change), or (3) a different object in the same location (identity change). For a detailed illustration of the timing of the experimental phases, see Figure 1 in Thiele et al. (19). For an example video illustrating the temporal sequence of a single trial, see Movie S1 in the Supplementary Materials.

During the first testing session, dogs participated in two pretest trials before the main experiment started. These trials were designed to familiarize the dogs with the depicted actors, the structured sequence of action, occlusion, and outcome phases used in the test trials, and the two possible object positions (19–21, 25). The pretest videos followed the same temporal structure but without referential gaze shifts during the action phase. Instead, the actors turned either toward or away from the dog. The outcome phase always showed the same object in the same location. The actor in the eye-contact pretest was the same as in the eye-contact test trials, and likewise for the no-eye-contact condition.

Each dog saw two different actors, one for each condition, and the assignment of actors to conditions was counterbalanced across subjects. Sixteen toy images (260x260 px) were used during test trials, counterbalanced across the six experimental conditions (2 eye contact × 3 outcome types), such that each object served both as a familiar and a novel item across participants. Two additional objects were used in the pretest phase. The toys were selected from a validated set used in prior infant studies (14, 19), ensuring comparability with human data. Additionally, we matched the luminance of the object images by scaling the RGB values of non-transparent pixels to a common target value. We selected visually distinct toys that included common features found in typical dog toys (e.g., rope texture, plush material, rubber shapes). Object screen positions, either top or bottom, were also counterbalanced within and across dogs and conditions to ensure balanced exposure. In the first session, outcome types and social conditions were fully counterbalanced across dogs. In subsequent sessions, trial orders were pseudo-randomized with constraints to prevent more than two consecutive trials of the same type and to ensure all three outcomes appeared once per session.

Procedure

We used an EyeLink 1000 eye-tracking system (SR Research, Canada) to examine dogs' gaze behavior in response to the video stimuli. The dogs were trained to place their heads on a chinrest before participating in the experiment; details about the training are provided in Karl et al. (56). During the training and before each test run, each dog performed 5-point calibrations with animated, non-social targets (size: 64 x 64 px; e.g., a moving fly or cartoon figure). The training was considered successful if dogs achieved a calibration validation with an average of less than 1° of visual angle deviation and maintained their head on the chinrest for at least one minute when the trainer left the room. During training and testing, each dog rested its head on an adjustable chinrest tailored to their body size to ensure a stable head position and consistent viewing distance (approximately 67 cm) from the monitor. Videos played at 25 fps on a 1920 × 1080 px LCD monitor, covering 43.4° by 25.2° of visual angle. Interested owners were allowed to remain in the room during the experiments and watch the screen from behind their dog's back, approximately two meters from the dog.

At the beginning of each trial, an attention-grabbing animation (identical to the one used during the calibration and different from the one used in the human infant study by Thiele et al. (19)) was presented at the center of the screen to direct and center the dogs' gaze. The animation preceding the action phase was gaze-contingent, meaning that the trial would only begin if the dog's gaze were detected for 50 msec within a predefined area of interest (AoI) surrounding the animation (concentric AoI, diameter = 300 px). This procedure ensured engagement and reduced data loss due to inattention. After each video presentation, a gray screen appeared until the following fixation animation or the end of the experiment. If a trial needed to be terminated before the video concluded, such as when the dogs left the chinrest, that trial, and any subsequent trials, were repeated after recalibrating, either in the same session or in a subsequent one (this happened in six trials). During the whole session, water was available to the dogs.

Statistical analysis

We fitted Generalized Linear Mixed Models (GLMM) with gamma error structure and Linear Mixed Models (LMM) to analyze the dogs' looking times into the object AoI (AoI size: w x h: 350 x 350 px), latency to look at the object AOI, and first look duration. Following our preregistered analysis plan, we fitted gamma-distributed GLMMs unless the model assumptions were violated. We checked whether the model assumptions were violated using the 'plotQQunif' (QQ-plot of residuals) and 'plotResiduals' (residuals plotted against predicted values) functions of the DHARMa package (57). Significant violations concerning the distribution of the residuals were detected for the looking time model, but not for the first look duration model. Therefore, we fitted an LMM for the overall dwell time data instead (and evaluated the Gamma model in the case of the first look duration). We analyzed the overall dwell times as within-subject z-scores to account for individual differences in absolute looking times and to ensure that the assumptions of the LMM were met (e.g., 57). Following the z-transformation, we found no apparent deviations from model assumptions concerning normal distribution and homoscedasticity of residuals. Our primary analyses were based on fixations as determined by the default Eyelink event parsing algorithm, but we also repeated our confirmatory analyses based on the raw samples (excluding blink artefacts; see ESM).

As predictor variables, we include the eye-contact condition (eye-contact or no-eye-contact), outcome type (no-change, identity-change, location-change) and their interaction, as well as the object position in the outcome phase (top or bottom) and overall trial number (z-transformed). The session number and the number of trials per session varied across dogs. Therefore, we included the overall trial number instead of the session number and trials within the session. Dog ID was included as a random effect as well as (centered) random slope components of eye-contact condition, outcome, object position, and trial number to avoid pseudo-

replication and to keep the type-I error rate at the nominal level of 5% (59). In the case of convergence issues, we removed the correlation between random intercepts and random slopes. We evaluated the significance of the interaction by using a likelihood ratio test. Following a significant interaction, we tested the difference between the identity-change vs no-change and location-change vs no-change for each condition of individual fixed effects employing the Satterthwaite approximation using the function lmer of the package lmerTest and a model fitted with restricted maximum likelihood.

For the exploratory analysis, including latency to first fixation within the object AoI (outcome phase), looking time to the object (action phase), and looking time to the actor (action phase; AoI size: w x h: 550 x 720 px), we fitted LMMs based on within-subject z-scores. The model structure was the same as in the confirmatory models (except that the outcome condition was, of course, not added for the action phase models).

All statistical analyses were conducted in R (59; version 4.4.2) using the packages glmmTMB (version 1.1.10; 59), lme4 (version 1.1-35.5; 60), and lmerTest (63). We obtained confidence intervals for model estimates and fitted values using a parametric bootstrap (N = 1,000 bootstraps). We assessed potential collinearity using Variance Inflation Factors (VIF; 62), which were calculated for a linear mixed model without the interaction term and random effects. The results indicated that collinearity was not a concern, with the maximum VIF being <1.1.

Thirty-six dogs started the experiment, but only 34 completed all test sessions. One additional dog completed the first test session (included); another dog began with the pretest but became unavailable before the start of the test phase (not included). Out of the 409 recorded test trials, we excluded trials if the dogs' gaze was detected in less than 70% of the action phase video (N=8) and if subjects looked at the object for less than 100 msec in the outcome phase (N=115). Given the dogs' bias to look at the upper half of the screen and the fact that there was no gaze-contingent onset of the outcome phase, we also filtered trials in which the dogs were looking at the object's outcome location at the start of the outcome phase and, therefore, were not actively shifting their gaze to the object once it appeared (N = 39). The final sample consisted of a total of 247 trials of 35 dogs.

Data, Materials, and Software Availability

All data, stimuli, and analysis scripts used in this study are publicly available on Zenodo (https://zenodo.org/records/17553617). The repository contains raw and pre-processed eye-tracking data, stimulus materials and all R scripts required to replicate the pre-processing, modelling and statistical analyses described in the manuscript. There are no restrictions on data access.

Acknowledgments

We want to thank the dogs and their caregivers, Laura Laussegger and Marion Umek, for collecting the data, and Karin Bayer for providing administrative support. This research was funded in whole or in part by the Austrian Science Fund (FWF) [grant DOI: 10.55776/P36896]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

References

1. M. Tomasello, “Joint attention as social cognition” in *Joint Attention*, (Psychology Press, 2014), pp. 103–130.

2. M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* **28**, 675–691 (2005).

3. J. Barresi, C. Moore, Intentional relations and social understanding. *Behavioral and Brain Sciences* **19**, 107–122 (1996).

4. J. Bruner, Joint attention: Its origins and role in development. *(No Title)* (1995).

5. C. Moore, Understanding self and others in the second year. *Socioemotional development in the toddler years: Transitions and transformations* 43–65 (2007).

6. R. Bakeman, L. B. Adamson, Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development* 1278–1289 (1984).

7. M. Carpenter, K. Liebal, Joint attention, communication, and knowing together in infancy. (2011).

8. M. Abels, Triadic interaction and gestural communication: Hierarchical and child-centered interactions of rural and urban Gujarati (Indian) caregivers and 9-month-old infants. *Developmental Psychology* **56**, 1817 (2020).

9. M. Botero, Tactless scientists: Ignoring touch in the study of joint attention. *Philosophical Psychology* **29**, 1200–1214 (2016).

10. E. E. Little, L. J. Carver, C. H. Legare, Cultural variation in triadic infant–caregiver object exploration. *Child Development* **87**, 1130–1145 (2016).

11. M. Thiele, G. Gredebäck, D. Haun, M. Thiele, Social Connectedness Without Eye Contact: 18-But Not 9-Month-Olds Use Proximal Touch to Infer Third-Party Joint Attention During Observational Learning. (2025).

12. S. Hoehl, S. Wahl, S. Pauen, Disentangling the effects of an adult model's eye gaze and head orientation on young infants' processing of a previously attended object. *Infancy* **19**, 53–64 (2014).

13. V. M. Reid, T. Striano, Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *European Journal of Neuroscience* **21**, 1763–1766 (2005).

14. S. Wahl, C. Michel, S. Pauen, S. Hoehl, 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 (2013).

15. A. Cleveland, M. Schug, T. Striano, Joint attention and object learning in 5-and 7-month-old infants. *Infant and Child Development: An International Journal of Research and Practice* **16**, 295–306 (2007).

16. A. Cleveland, T. Striano, The effects of joint attention on object processing in 4-and 9-month-old infants. *Infant Behavior and Development* **30**, 499–504 (2007).

17. Y. Okumura, Y. Kanakogi, T. Kobayashi, S. Itakura, Ostension affects infant learning more than attention. *Cognition* **195**, 104082 (2020).

18. M. Thiele, R. Hepach, C. Michel, D. Haun, Observing others' joint attention increases 9-month-old infants' object encoding. *Developmental Psychology* **57**, 837 (2021).

19. M. Thiele, S. Kalinke, C. Michel, D. B. M. Haun, Direct and Observed Joint Attention Modulate 9-Month-Old Infants' Object Encoding. *Open Mind* **7**, 917–946 (2023).

20. Y. Okumura, T. Kobayashi, S. Itakura, Eye Contact Affects Object Representation in 9-Month-Old Infants. *PLOS ONE* **11**, e0165145 (2016).

21. J. M. Yoon, M. H. Johnson, G. Csibra, Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences* **105**, 13690–13695 (2008).

22. G. Csibra, G. Gergely, Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI* **21**, 249–274 (2006).

23. G. Csibra, G. Gergely, Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 1149–1157 (2011).

24. G. Csibra, R. Shamsudheen, Nonverbal Generics: Human Infants Interpret Objects as Symbols of Object Kinds. *Annual Review of Psychology* **66**, 689–710 (2015).

25. P. Silverstein, T. Gliga, G. Westermann, E. Parise, Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008). *Infant Behavior and Development* **55**, 77–87 (2019).

26. H. Marno, E. J. Davelaar, G. Csibra, Nonverbal communicative signals modulate attention to object properties. *Journal of Experimental Psychology: Human Perception and Performance* **40**, 752–762 (2014).

27. H. Marno, E. J. Davelaar, G. Csibra, An object memory bias induced by communicative reference. *Acta Psychologica* **163**, 88–96 (2016).

28. H. Marno, C. J. Völter, B. Tinklenberg, D. Sperber, J. Call, Learning from communication versus observation in great apes. *Sci Rep* **12**, 2917 (2022).

29. A. Catala, B. Mang, L. Wallis, L. Huber, Dogs demonstrate perspective taking based on geometrical gaze following in a Guesser–Knower task. *Anim Cogn* **20**, 581–589 (2017).

30. L. Huber, P. van der Wolf, M. Menkveld, S. Riemer, C. J. Völter, Canine perspective taking: Anticipating the behavior of an unseen human. *iScience* **28** (2025).

31. J. Kaminski, J. Brauer, J. Call, M. Tomasello, Domestic dogs are sensitive to a human's perspective. *Behaviour* **146**, 979–998 (2009).

32. A. Kis, *et al.*, Does the A-not-B error in adult pet dogs indicate sensitivity to human communication? *Animal Cognition* **15**, 737–743 (2012).

33. L. Lonardo, C. J. Völter, C. Lamm, L. Huber, Dogs follow human misleading suggestions more often when the informant has a false belief. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20210906 (2021).

34. L. Lonardo, M. Putnik, V. Szewczak, L. Huber, C. J. Völter, Dogs do not use their own experience with novel barriers to infer others' visual access. *Proceedings of the Royal Society B: Biological Sciences* **291**, 20232934 (2024).

35. M. E. Maginnity, R. C. Grace, Visual perspective taking by dogs (Canis familiaris) in a Guesser–Knower task: evidence for a canine theory of mind? *Anim Cogn* **17**, 1375–1392 (2014).

36. E. Téglás, A. Gergely, K. Kupán, A. Miklósi, J. Topál, Dogs' gaze following is tuned to human communicative signals. *Curr. Biol.* **22**, 209–212 (2012).

37. J. Topál, G. Gergely, Á. Erdőhegyi, G. Csibra, Á. Miklósi, Differential sensitivity to human communication in dogs, wolves, and human infants. *Science* **325**, 1269–1272 (2009).

38. Zs. Virányi, J. Topál, Á. Miklósi, V. Csányi, A nonverbal test of knowledge attribution: a comparative study on dogs and children. *Anim Cogn* **9**, 13–26 (2006).

39. C. J. Völter, K. Gerwisch, P. Berg, Z. Virányi, L. Huber, Using mobile eye tracking to study dogs' understanding of human referential communication. *Proceedings of the Royal Society B: Biological Sciences* **292**, 20242765 (2025).

40. J. Call, J. Bräuer, J. Kaminski, M. Tomasello, Domestic dogs (Canis familiaris) are sensitive to the attentional state of humans. *Journal of comparative psychology* **117**, 257 (2003).

41. C. Schwab, L. Huber, Obey or not obey? Dogs (Canis familiaris) behave differently in response to attentional states of their owners. *Journal of Comparative Psychology* **120**, 169 (2006).

42. J. Kaminski, A. Pitsch, M. Tomasello, Dogs steal in the dark. *Animal cognition* **16**, 385–394 (2013).

43. K. Kupán, Á. Miklósi, G. Gergely, J. Topál, Why do dogs (Canis familiaris) select the empty container in an observational learning task? *Animal Cognition* **14**, 259–268 (2011).

44. Z. Sümegi, A. Kis, Á. Miklósi, J. Topál, Why do adult dogs (Canis familiaris) commit the A-not-B search error? *Journal of Comparative Psychology* **128**, 21 (2014).

45. T. Tauzin, A. Csík, A. Kis, J. Topál, What or where? The meaning of referential human pointing for dogs (Canis familiaris). *Journal of Comparative Psychology* **129**, 334–338 (2015).

46. A. Bergt, A. E. Urai, T. H. Donner, L. Schwabe, Reading memory formation from the eyes. *European Journal of Neuroscience* **47**, 1525–1533 (2018).

47. A. Kafkas, D. Montaldi, Recognition Memory Strength is Predicted by Pupillary Responses at Encoding While Fixation Patterns Distinguish Recollection from Familiarity. *Quarterly Journal of Experimental Psychology* **64**, 1971–1989 (2011).

48. M. Ishikawa, S. Itakura, Physiological arousal predicts gaze following in infants. *Proc Biol Sci* **286**, 20182746 (2019).

49. W. Pätzold, U. Liszkowski, Pupillometry reveals communication-induced object expectations in 12- but not 8-month-old infants. *Dev Sci* **22**, e12832 (2019).

50. B. Hare, M. Brown, C. Williamson, M. Tomasello, The domestication of social cognition in dogs. *Science* **298**, 1634–1636 (2002).

51. S. Karl, *et al.*, Exploring the dog–human relationship by combining fMRI, eye-tracking and behavioural measures. *Sci Rep* **10**, 22273 (2020).

52. F. Range, U. Aust, M. Steurer, L. Huber, Visual categorization of natural stimuli by domestic dogs. *Anim Cogn* **11**, 339–347 (2008).

53. C. J. Völter, S. Karl, L. Huber, Dogs accurately track a moving object on a screen and anticipate its destination. *Sci Rep* **10**, 19832 (2020).

54. C. J. Völter, A. Tomašić, L. Nipperdey, L. Huber, Dogs' expectations about occlusion events: From expectancy violation to exploration. *Proceedings of the Royal Society B* **290**, 20230696 (2023).

55. C. J. Völter, L. Huber, Dogs' looking times and pupil dilation response reveal expectations about contact causality. *Biology Letters* **17**, 20210465 (2021).

56. S. Karl, M. Boch, Z. Virányi, C. Lamm, L. Huber, Training pet dogs for eye-tracking and awake fMRI. *Behav Res* **52**, 838–856 (2020).

57. F. Hartig, DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *CRAN: Contributed Packages* (2016).

58. F.-S. Koch, *et al.*, Procedural memory in infancy: Evidence from implicit sequence learning in an eye-tracking paradigm. *Journal of Experimental Child Psychology* **191**, 104733 (2020).

59. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278 (2013).

60. R. C. Team, R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. (2019).

61. M. Brooks, *et al.*, glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* (2017).

62. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

63. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1–26 (2017).

64. A. P. Field, Discovering statistics using SPSS:(and sex, drugs and rock'n'roll). *(No Title)* xxxiv (2005).

Figures and Tables.

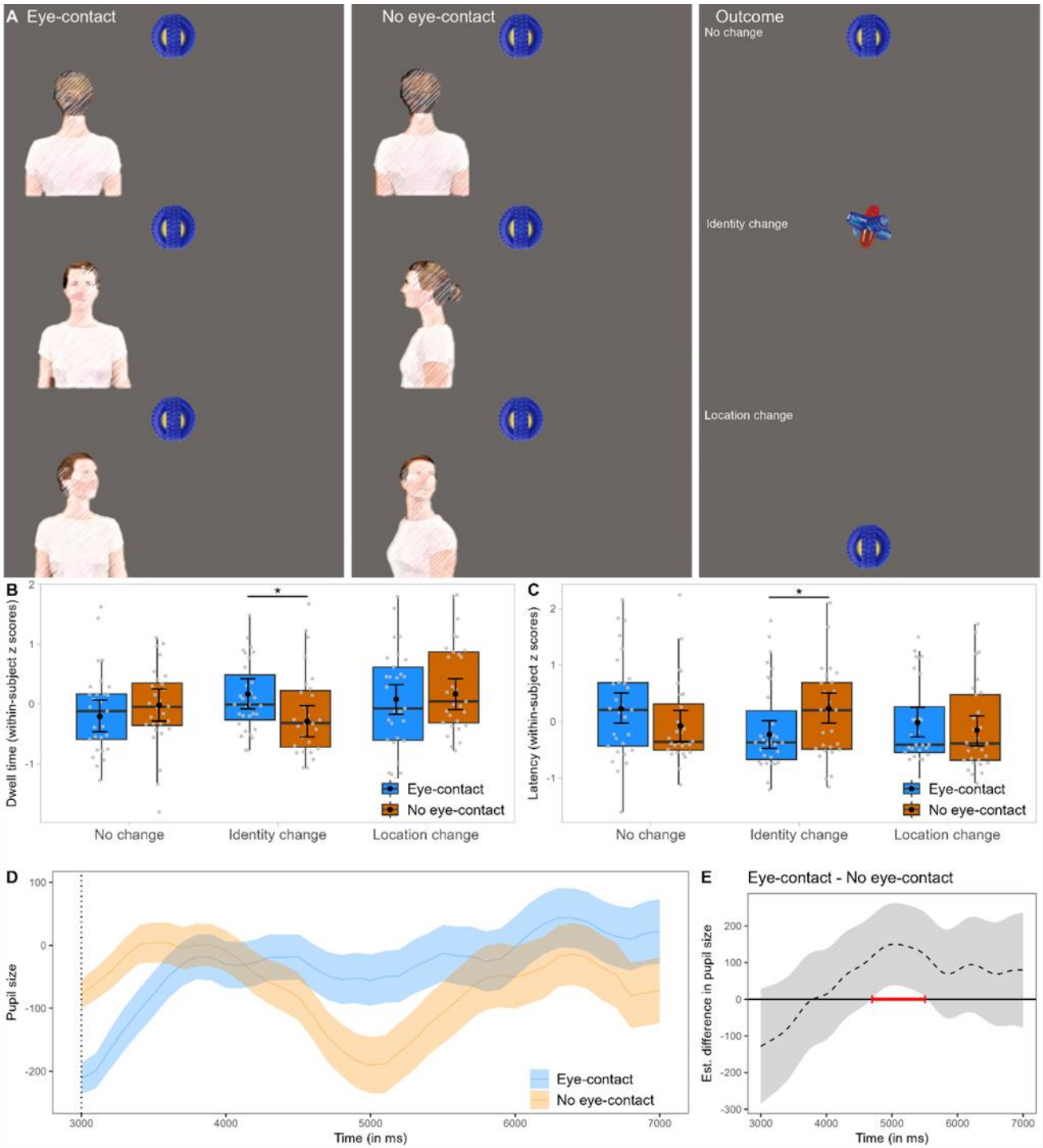

Figure 1. Screenshots and data visualizations. (A) Screenshots of the video stimuli in the eye-contact (left) and no-eye-contact (center) conditions of the still phases in the action phase and of the outcome conditions in the outcome phase (right). The positioning of the actor (left or right on screen) and of the object in the action phase (top or bottom) was counterbalanced. (B) Boxplot showing the dogs' dwell time (within-subject

z-scores) in the object AoI during the outcome phase across the eye-contact and outcome conditions. (C) Latency (shown as within-subject z-scores) of the first look into the object AoI in the outcome phase. In (B) and (C), data are grouped by condition, with eye-contact trials in blue and no-eye-contact trials in orange. Grey dots represent the mean values of each subject, while black dots show the fitted model, and the error bars are its 95% confidence interval: *$p < 0.05$. (D) Time series plot showing dogs' pupil size (in arbitrary units and baseline corrected). The blue and orange lines show the mean pupil size (± SE, shaded area around the line) in the action phase of the eye-contact and no-eye-contact conditions. The dashed vertical line (at 3000 ms) indicates when the actor turns toward the camera in the eye-contact condition. (E) Difference curve obtained from GAMM01. The dashed line shows the estimated difference between the eye-contact and no-eye-contact conditions during the action phase, and the shaded area shows the pointwise 95% confidence interval. The period in which the conditions differ significantly is highlighted in red. *Based on stimuli originally developed for Thiele, M., Kalinke, S., Michel, C., & Haun, D. B. (2023). Direct and observed joint attention modulate 9-month-old infants' object encoding. Open Mind, 7, 917–946.*