



ISPA

INSTITUTO UNIVERSITÁRIO
CIÊNCIAS PSICOLÓGICAS, SOCIAIS E DA VIDA

**Do dogs (*Canis familiaris*)
predict a conspecific's next
action based on its gaze?**

Catarina Luísa Espanca Bacelar

Orientador de Dissertação:

Dr. Zsófia Virányi

Professor de Seminário de Dissertação:

Professor Doutor Rui Oliveira

Tese submetida como requisito parcial para a obtenção do grau de:
MESTRE EM NEUROCIÊNCIAS COGNITIVAS E COMPORTAMENTAIS

2017

Dissertação de Mestrado realizada sob a orientação de
Dr. Zsófia Virányi, apresentada no ISPA – Instituto
Universitário para obtenção de grau de Mestre em
Neurociências Cognitivas e Comportamentais

Acknowledgements

Special thanks to Dr. Zsófia Virányi for allowing my dream of working at the Clever Dog Lab to become reality and for her amazing help especially in the final countdown and to Soon Young Park for flooding me with information every time that I had a doubt and for teaching me about eye-tracking and statistics and to help me construct my experiment. To Dr. Tamás Faragó for all the useful discussions of how to make the experiment. To Karin Bayer for receiving me with open arms since the first day until the last one. To Sabrina Karl for training the dogs so calmly well. To Jenny Bentlage for her Michel, the actor dog. To Wolfgang Berger for building a big part of the experimental equipment. Finally, to all the owners and their dogs who participated in this study.

“Conseguem os cães (*Canis familiaris*) prever a próxima ação de outro cão através do seu olhar?”

Resumo

Para onde um indivíduo olha é uma das pistas mais confiáveis para prever a sua próxima ação. Neste estudo investigamos se os cães domésticos através do olhar de outro cão conseguem prever qual de duas taças este terá maior probabilidade de se aproximar. Usámos um paradigma de violação de expectativas, onde o tempo de atenção visual dos participantes nos diz se estes criaram alguma expectativa em relação a onde o outro indivíduo irá procurar comida depois de terem visto esse indivíduo a olhar para uma das duas taças. O tempo de atenção visual foi medido com um sistema de *eye-tracking*. Na Experiência 1, os participantes viram dois tipos de vídeos. Na condição congruente, o ator olhou para uma das duas taças e depois comeu da mesma taça. Na condição incongruente, o ator olhou para uma taça mas depois aparece a comer da outra taça. Resultados mostraram que os participantes olharam mais tempo para o ator a comer no primeiro teste incongruente do que no primeiro teste congruente, enquanto que têm uma tendência para olhar mais tempo para a taça onde o ator estava a comer no teste congruente do que no teste incongruente. Na Experiência 2, os testes congruentes e incongruentes foram comparados com controlos onde o ator não forneceu nenhuma pista visual antes de comer de uma das taças. Aqui não obtemos diferenças entre testes e controlos. Os resultados sugerem que os cães conseguem usar o olhar de outro cão para prever a sua próxima ação e possivelmente podem também reconhecer o conteúdo referencial do olhar.

Palavras-chave

Seguir o olhar do outro, compreensão da ação, violação de expectativas, *Canis familiaris*

“Do dogs (*Canis familiaris*) predict a conspecific’s next action based on its gaze?”

Abstract

Where someone is looking is one of the most reliable cue predicting her/his next action. Here we set out to investigate whether pet dogs predict which of two objects another dog is likely to approach based on its gaze direction. We used a violation-of-expectation paradigm, where the dogs’ looking time can tell if they created an expectation where another dog will look for food after they have seen it watching one of two food locations. Looking times were measured with an eye-tracker. In Experiment 1, subjects saw two kinds of video clips. In the congruent condition, the actor dog looked at one of two bowls and then ate from the same bowl. In the incongruent condition, the actor looked at one bowl, but then was visible eating from the other bowl. We found that subjects looked longer at actor eating in the first incongruent trial than in the first congruent trial whereas they tended to watch the bowl the actor was eating from longer in the congruent than in the incongruent trial. In Experiment 2, similar test trials were compared to control trials where the actor provided no gaze cue before eating from a bowl. Here we found no differences between test and control trials. Findings suggest that dogs can use gaze to predict a conspecific's next action and they may also recognize the referential content of gaze.

Key-words

Gaze-following, action understanding, violation-of-expectation, *Canis familiaris*

INDEX

1. Introduction	1
2. Experiment 1	5
2.1. Methods	5
2.1.1. Subjects.....	5
2.1.2. Training	5
2.1.3. Eye-tracking system	6
2.1.4. Experimental setup	7
2.1.5. Stimuli	7
2.1.6. Experimental procedure.....	8
2.1.7. Data preparation	10
2.1.8. Data analysis.....	11
2.2. Results	12
2.3. Discussion.....	14
3. Experiment 2	15
3.1. Methods	16
3.1.1. Subjects.....	16
3.1.2. Stimuli	16
3.1.3. Experimental procedure.....	16
3.1.4. Data preparation	17
3.1.5. Data analysis.....	17
3.2. Results	18
3.3. Discussion.....	20
4. General Discussion	20
5. References	24
ANNEXES	A

LIST OF FIGURES

Figure 1. Test videos timeline	9
Figure 2. Chosen pot and body areas.....	12
Figure 3. Total durations of all fixations in Experiment 1	12
Figure 4. Proportion of durations of fixations on actor's body area in Experiment 1.....	13
Figure 5. Proportion of durations of fixations on chosen pot area in Experiment 1	14
Figure 6: Control videos timeline.....	17
Figure 7. Total durations of all fixations in Experiment 2	18
Figure 8. Proportion of durations of fixations on actor's body area in Experiment 2.....	19
Figure 9. Proportion of durations of fixations on chosen pot area in Experiment 2	19
Figure A1. Top view of the experimental room	A
Figure A2: General view of the experiment compartment with subjects in the chin-rest.....	A

1. Introduction

Members of social species need to coordinate their actions with the actions of others, and this often requires the prediction of others' actions. The ability to distinguish between physical and psychological objects and predict the latter's actions on other objects represents a highly adaptive capacity (Premack & Premack, 1997; Santos & Hauser, 1999). An interesting question is how far this capacity extends to the non-human animal world, and which cues animals rely on to predict others' action. Gaze direction has been suggested to be one of the most reliable cues preceding one's next action: individuals look where they attend and next intend to act (Shepherd, 2010).

Paying attention to others' gaze and the capacity to follow it have been suggested to play an important role in the development of social cognition (Fitch, Huber, & Bugnyar, 2010). For humans, detecting the direction of others' gaze is a central aspect of social interactions and communication (Brooks & Meltzoff, 2002). Human adults use the gaze of another individual as a deictic ("pointing") cue to see what the other is seeing, to predict the other's future actions and to attribute mental states (Shepherd, 2010). The capacities to coordinate with others' head and eye orientation to look in the same direction (gaze following) or at a specific target (joint attention) are considered a decisive step toward an understanding of mental states like attention and intention of others (Baron-Cohen, 1995; Tomasello, Carpenter, Call, Behne, & Moll, 2005), and have been implicated in the development of the so-called theory of mind (Baron-Cohen, 1995; Shepherd, 2010). Attributing mental states to others, that is inferring what kind of beliefs, desires, intentions and perspectives they have, is the most efficient way of predicting their future actions (Goldman, 2012). The development of these capacities take place early during ontogeny (Butterworth & Jarrett, 1991), and following others' gaze, via joint attention, is thought to facilitate also language acquisition in human children (Brooks & Meltzoff, 2002).

Phillips et al. (2002) investigated infants' ability to connect gaze and emotional expression to action. Their question was whether and when infants recognize that a person's emotional expression and visual regard connect to their actions and thus, can predict the direction of their movements. The authors term as "object-directedness" the fact that intentional actions on an object can manifest a distinctive directedness to that object, reflecting the intention's underlying goal-directedness. Specifically, when an agent acts intentionally toward an object the movement dynamics often result in an approach to (or avoidance of), contact with, acquisition of, and maybe even consumption of the object. They propose that the object-

directedness of behavior and the functional connections between the target action and other aspects of the person, especially their visual-emotional regard and expressions, generally manifest and potentially identify intentions. Therefore, exploring infants' use of the object-directedness and functional connections manifested in human action is of special interest. Thus, Phillips et al. (2002) focused on infants' recognition of connections between actors' actions and their perceptual-emotional displays. The authors tested 8 and 12-months-old infants using a violation-of-expectation paradigm. A violation-of-expectation paradigm has long been used to study infant cognition as it requires no complex behaviours from the infants. Instead it simply relies on the infants' tendency to look longer at unexpected and surprising events compared to familiar and expected events. A large body of evidence that subjects look reliably longer at unexpected than at expected events is thought to indicate that they possess the expectation under investigation, detect the violation in the unexpected event and are surprised (state of heightened attention or interest) by this violation (Wang, Baillargeon, & Brueckner, 2004). Infants saw two kinds of events, one consistent and one inconsistent with the principle that emotional and visual regard predict the directedness of a following action, specifically, if a person looks at and emotes positively about one object, rather than another, then he or she will be likely to grasp that first object rather than the other. In Phillips et al. (2002) study, the consistent event showed to the infants depicted an actor first looking at one of two stuffed kitties and showing positive emotions, and then the same actor holding this object. The inconsistent event also starts with the actor looking and emoting about one of the kitties, but then the actor appears holding the other object. Results showed that 12-month-olds, but not 8-month-olds, recognized that the actor was likely to grasp the kitty which she had visually regarded with positive affect, indicating that the ability to use information about an adult's direction of gaze and emotional expression to predict action develops in humans between the 8th and the 12th month of age.

The finding that 12-months infants are capable to predict actions based on an actor's gaze-direction was further supported by Sodian & Thoermer (2004). They used not only gaze but also different referential gestures, such as pointing, as well as goal-directed reaching and grasping actions to indicate the actor's interest in an object, and found that 12-month-old infants use also reaching and grasping movements as cues to action goals, whereas relying on pointing as a cue to action intention seems to develop later and follow a different route of acquisition.

Recognizing the referential aspect of gaze is obviously not the most limiting component of predicting others' next object-directed actions based on their gaze, as Woodward (2003) has shown that human infants can recognize that gaze establishes a relation between the looker and the object of his or her gaze. Her results indicate that infants begin to understand this relation

between 9 and 12 months of age. Also younger infants follow others' gaze but before this age gaze-following does not rely on an understanding of the relation between a person looking and the object of his or her gaze.

Although much is known about the development of human infants' capacity to use gaze as a behavioral predictor, and much less research has been conducted on this matter in non-human animals. Most relevant for the current thesis, Santos and Hauser (1999) investigated whether cotton-top tamarins (*Saguinus oedipus oedipus*) use the direction of a human's gaze to predict her/his future actions. They used a violation-of-expectation paradigm, similar to the one by Phillips et al. (2002). Subjects looked significantly longer at the inconsistent test event in which the actor reached for the one of two objects that she had not previously oriented to. However, there was no difference in looking time between consistent and inconsistent test events when the actor used only her eyes to orient. The authors have argued that the findings suggest that tamarins are able to use head orientation and gaze direction, but not gaze direction alone, to predict the actions of a human agent. Interestingly, another study testing squirrel monkeys (*Saimiri sciureus*) and capuchin monkeys (*Sapajus apella*) in a similar study found negative results (Anderson, Kuroshima, Kuwahata, & Fujita, 2004).

Given such scarce and contradictory results, this question calls for further investigation in non-human animals. As argued at the beginning of the introduction, it is likely that social species that readily use visual information do use gaze to predict others' future actions. In the present study, we intend to address this question in a non-primate species, specifically, we investigate whether pet dogs (*Canis familiaris*) use the gaze of another dog to predict his next action.

Dogs originated from wolves – social animals that engage in a number of cooperative behaviors, such as hunting, territory defense and taking care of their offspring, and thus may have evolved cognitive abilities that help them to predict and interpret the actions of other animals (Cooper et al., 2003). Alternatively or additionally, it has been suggested that the long-term association of dogs with humans has resulted, through a process of convergent evolution, in advanced, human-like, socio-cognitive skills in dogs (Hare & Tomasello, 2005; Miklósi et al., 2003). Supporting this expectation, Horowitz (2009) has suggested that dogs are sensitive to the visual attention of their play partner when engaging in dyadic rough-and-tumble play. Results of this study have shown that dogs use visual cues in communication (play signals were sent nearly exclusively to forward-facing dogs, while attention-getting behaviors were used most often when a playmate was facing away) and use flexible means to acquire attention (stronger attention-getters were used when a playmate was looking away or distracted, less

strong ones when the partner was facing forward or laterally). Horowitz (2009) proposes the existence of a “rudimentary theory of mind” in non-human animals, that is, the existence of precursors skills to theory of mind, and suggests the use of visual attention-grabbing cues as a good candidate to be one of those precursors.

In line with their use of visual communication, dogs have also been found to be sensitive to others’ gaze. Studies have found that dogs are sensitive to humans’ visual attention (e.g. Call, Bräuer, Kaminski, & Tomasello, 2003; Cooper et al., 2003; Gácsi, Miklósi, Varga, Topál, & Csányi, 2004) as well to conspecifics’ gaze cues (Hare & Tomasello, 1999; Herendy, 2011; Werhahn, Virányi, Barrera, Sommese, & Range, 2016).

However, until now, to our knowledge, no study has directly addressed the question whether dogs use conspecific gaze to predict another dog's next action. Here, we explore this question by adapting the violation-of-expectation paradigm used by Santos and Hauser (1999) to dogs. Violation-of-expectation paradigms have already been used successfully in other dog studies (Müller, Mayer, Dörrenberg, Huber, & Range, 2011; West & Young, 2002) but in this instance an eye-tracker system is used to measure how long the dogs pay attention to which part of the video stimuli used.

Here we presented the dogs with an object-choice situation, a context where dogs have been found to follow the gaze of conspecifics (Hare & Tomasello, 1999; Herendy, 2011). The subjects watched a video clip in which an actor dog, after facing the subjects while sitting in between two identical food containers, turns his head and looks at one of the containers. After this gaze cue the screen turned black, and then we presented subjects with two kinds of events, a congruent and an incongruent. In the congruent event, the actor dog ate from the pot that he had looked at (expected outcome), whereas in the incongruent event, the actor ate from the other pot (unexpected outcome). We hypothesized that if dogs are sensitive to the link between the direction of actor’s gaze and his subsequent action, they should look significantly longer to the incongruent event. In addition, we were interested if the subject watches mainly the actor dog or pay also attention to the food container, suggesting that they recognize the target object of his actions.

Subjects’ looking behaviors were recorded with an eye-tracker system and the event presentations were made through video projection since live presentations would differ greatly from each other (due to several procedure difficulties associated with the live reproduction of such a presentation) and would make the eye-tracking recording more difficult. Studies have already shown that dogs can reliably use projected images and pre-recorded videos successfully (Péter, Miklósi, & Pongrácz, 2013; Somppi, Törnqvist, Hänninen, Krause, & Vainio, 2012).

To avoid possible novelty and familiarity preferences induced by the habituation or familiarization trials included in the original violation-of-expectation paradigm by Phillips et al. (2002), we conducted only test trials without familiarization trials beforehand. Other studies have already demonstrated that violation-of-expectation tasks that include only test trials can provide evidence by means of violations of expectations (Müller et al., 2011; Wang et al., 2004).

As the aim of this study was to compare the looking time between congruent and incongruent outcomes, we confined the analyses to the outcome phase, as the phases before would not include the response to the outcome.

2. Experiment 1

2.1. Methods

All experiments were conducted at the Clever Dog Lab, Vetmeduni Vienna, Austria from September 2015 to January 2016. Procedures were approved by the Ethical Committee for the Use of Animals in Experiments at the Vetmeduni.

2.1.1. Subjects

The dogs were all pets volunteered by their owners to participate in this study. They were recruited through the Clever Dog Lab dog database on a voluntary basis. The 18 subjects were adults, 11 females ranging from 1 to 11 years old ($M=4.73$, $SD=2.76$) and 7 males ranging from 1 to 7 years old ($M=3.86$, $SD=2.27$). All dogs had already participated in other eye-tracking studies. Dogs were neither restrained nor forced to perform the tasks. Their owners were continuously present during their task, but were asked not to interact with their dogs during the experiments.

2.1.2. Training

In order to prepare the dogs for eye-tracking, prior to the experiments they received a training to ensure optimal eye-movement recording. This included training the dogs to keep their head motionless on a chin-rest and to look at small light spots on a screen. Looking at these spots was crucial to enable calibration and validation of the eye-tracker that were performed before the experiments in order to "teach" the eye-tracker how an individual dog's

eye looks when looking at known locations. Based on this information the eye-tracker could then calculate what the dogs were watching when showing them video clips.

Training was performed by a professional dog trainer and by the experimenters using operant-positive conditioning methods (e.g. shaping) with a clicker to mark the desired behavior and food to reward the dogs. At the beginning, looking at the light spots and staying in the chin-rest were trained separately, in two different rooms, and at a later stage they were integrated. The training took place following different phases:

- 1) Dogs were trained to either sit or stand and to lean their muzzle on the chin-rest and to remain in that position for increasing periods of time up to 60 seconds.
 - 2) In parallel, in another room, they were trained to touch a light spot moved around on a large screen with their nose. Later on, touching was removed and the dogs were rewarded for only looking at the spots without approaching the screen. The dogs moved to the next stage of training when they were able to fixate their gaze on the spot at least seven different positions consecutively, with no interval (no reward) in between.
 - 3) When the dogs were able to perform both behaviors reliably, the light spot training was integrated with using the chin-rest. That is, the dogs were trained to look at the spots on the screen while keeping their head on the chin-rest.
 - 4) Finally, in order to extend the time the dogs were willing to spend in the chin-rest without moving to a sufficient length (at least 20 sec), video clips were added to the training phase.
- After the dogs reached this criterion, they could participated in one of our 2 experiments. The duration of the training depended on the learning progress of each dog and the owners' availability to come to the lab. On average 15 training sessions were needed to learn the required behaviors (usually 1 session per week) (range: 8 to 30 sessions).

2.1.3. Eye-tracking system

The eye movement of subjects was recorded by a high-resolution eye-tracker, EyeLink 1000 (video-based recording system) from SR Research Ltd. The system includes a high-speed camera with an infrared illuminator attached to its right, a host computer where the EyeLink 1000 application runs and a display computer where the stimuli presentation is controlled. The system estimates the gaze location of subjects through the geometrical relationship of the corneal and pupil reflection. Specifically, for the gaze location estimation, the EyeLink1000 system identifies the geometrical relationship of the two centers of an eye, the pupil center and the center of the reflection circle of the infrared light on the cornea in every eye image, sampled by the camera at a 1000 Hz rate. If the detection of the two centers fails, due to blinking for

instance, the tracker recovers sampling in less than 1.8 milliseconds, and the fails are logged in the recording data as a sample with no pupil diameter or missing.

2.1.4. *Experimental setup*

The experiments were conducted in the Black Room of the Clever Dog Lab (size: 8.5x3.6m, height: 3.0m). The room was divided in 2 compartments, the experimental (3.6x6.0m) and the control compartment (3.6x2.5m). The two compartments were separated with a 2x2m screen that was used to present the stimuli, and two doors placed one on each side of the screen (ANNEX I: experimental set-up).

The experiment compartment was the place where subject and owner were during the experiment and contained an adjustable metal structure that supported the chin-rest for the subjects and the eye-tracking camera with the infrared illuminator attached. The camera was positioned about 50-55cm in front of the chin-rest so that it could record the right eye of a subject when positioned in the chin-rest. The screen was about 200cm away from the chin-rest. The heights of the chin-rest and camera were adjusted to the height of each subject. The compartment was dimly lit (approximately 80 lx.) by LED lamps on the ceiling (no windows). In the control compartment, behind the screen was a projector that rear projected the stimuli on the screen. The projector was connected to the display computer where the stimuli presentation was controlled which in turn was connected to the EyeLink1000 host computer. The projector was positioned 70cm away from the screen and at a height of 40cm. Subjects were monitored during experiments through the host computer monitor.

2.1.5. *Stimuli*

The stimuli presented to the subjects were two video clips, congruent and incongruent, depicting an object-choice situation where a dog (the actor) sat between two food containers, while either looking at one of the containers or simply watching straight ahead and then approached one of the containers and ate from it (Figure 1).

The actor dog, Michel, was a 10-year-old castrated mongrel male weighing 15 kg and with light brown, yellowish pelage with some white spots. Michel had previously been trained to sit between the food containers, to look at them and to approach them on command, and then these behaviors were recorded several times using a Canon 6D camera (resolution 1920x1080; file format .mov). One of the videos sets started with Michel sitting in the middle of two identical food pots and looking forward, in the direction of the subject for 2 secs (No Cue Phase). Then he turned his head towards one of the pots (Gaze Cue Phase). In another set of

video clips, he was standing next to one of the pots and was eating from it for 8 sec (Choice Phase). Of these 2 sets, 2 video clips were selected and then edited. The first video contained the No Cue Phase and the Gaze Cue Phase where Michel looked at the left pot. To extend the gaze cue time, the last frame where Michel was looking at the pot was frozen for 4 sec. In this way this clip had a final length of 6 sec. The second video clip showed the Choice Phase where Michel was eating from the left pot. In order to create the right gaze cue and the right choice video clips, we horizontally mirrored these 2 original clips. The video editing was made with VSDC Free Video Editor version 3.2.1.372 from Flash-Integro LLC.

Using these 4 clips we created 4 video stimuli, each lasting 18 sec. In each video after the No Cue (2 sec) and the Cue Phases (6 sec) there was a pause where no image appeared but the screen remained black for 2 sec which is about the time that Michel would have needed to approach one of the pots from his position in the middle. After this Black Pause Phase the Choice Phase (8 sec) followed (Figure 1).

In this way, combining gaze and choice to the same side, we had congruent videos: Congruent Left (CL) and Congruent Right (CR) where the gaze cue was consistent with the following action (choice), i.e. the pot that Michel approached was the one that he had looked at beforehand. In addition, combining gaze and choice to different sides we created two incongruent videos: Incongruent Left (IL) where Michel approached the left pot after he had looked at the right one and Incongruent Right (IR) where he approached the right pot after having looked at the left one.

Videos did not have sound and were rear projected in the middle of the 2x2m screen on a display area of 1,14m x 0,76 m. The stimuli presentation was designed and posteriorly controlled on Experiment Builder software from SR Research Ltd., a graphical programming environment for creating computer-based experiments.

2.1.6. Experimental procedure

Showing a video and recording the gazing pattern of a subject during video presentation was considered a trial. Each subject participated in 4 trials, each with one of the 4 test videos (CL, CR, IL and IR). The videos were presented in blocks of 2 trials, each of which contained a congruent and an incongruent trial. The order of conditions and sides of Michel's choice was counterbalanced across subjects. The 2 blocks were presented in one or two sessions, in case of the subject got tired or distracted before finishing the 4 trials.

In the beginning of each session, the trainer sent the subject to the chin-rest, and once its chin was in a stable position, the experimenter positioned the eye-tracking camera so as to

view the subject's right eye in the monitor of the host computer and adjust the thresholds of the system to make the detection of pupil and corneal reflection optimal. Then the experimenter as well as the trainer moved to the control compartment and closed the doors. The owner stayed behind the dog and was asked to remain still and silent during calibration and stimuli presentation. The experimenter verified on the control screen if the subject's eye was being captured correctly by the eye-tracker and then started the calibration after which the subjects directly proceeded with the stimuli presentation.

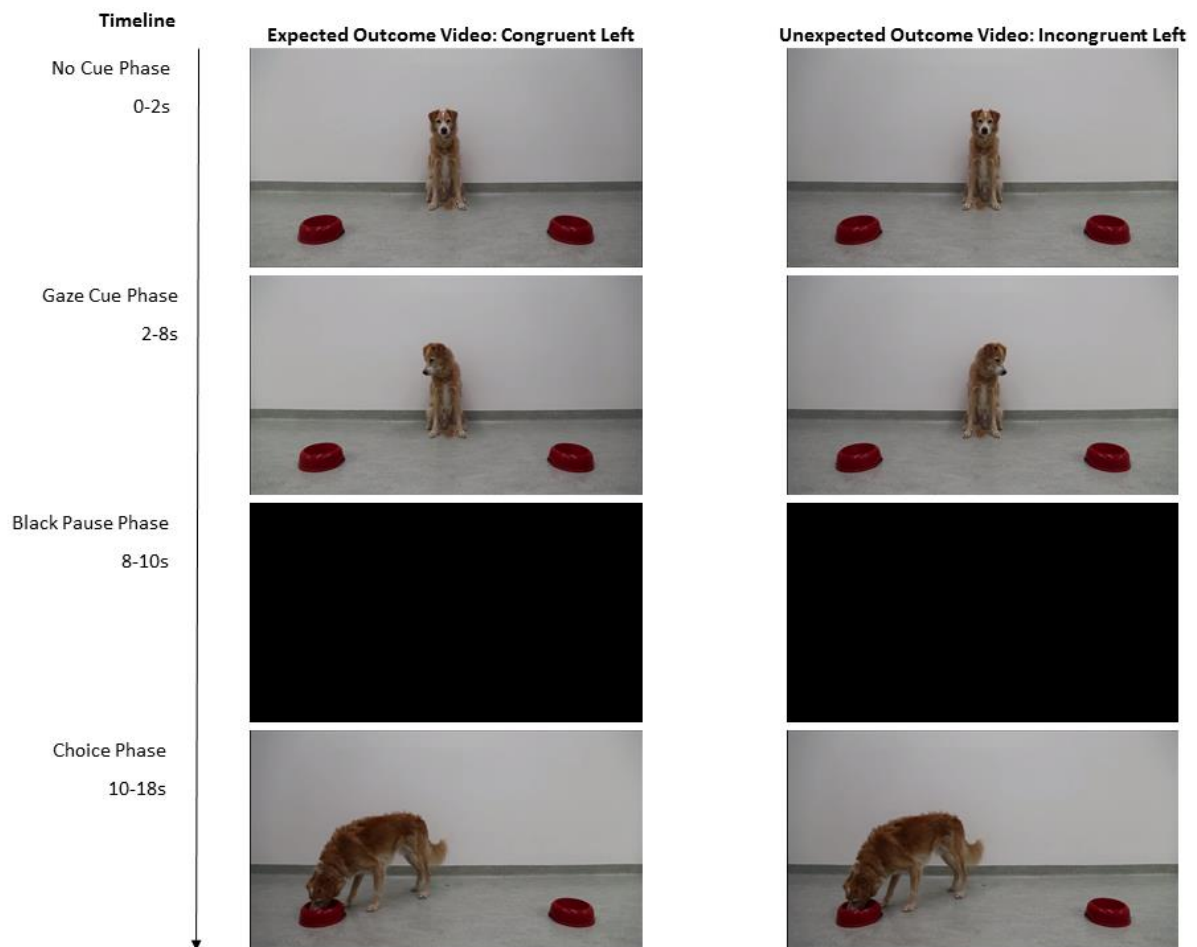


Figure 1. Test videos timeline - two of the 4 test videos showing the timeline of the 4 phases: No Cue, Gaze Cue, Black Pause and Choice Phase. On the left, screenshots from Congruent Left video and on the right, screenshots from Incongruent Left video.

Calibration of the eye-tracker

As mentioned, each experiment starts with the calibration of the eye-tracking system, a procedure that allows the system to know where the dog is looking at. This procedure consists in showing on the screen points of light that the subject has to fixate. The light points appear in 3 predefined places, which together constitute the vertices of an imaginary isosceles triangle.

The light points appear one at a time and the experimenter can visually check in the monitor of the host computer if the subject is fixating the point or not. Once the subject fixates a point, the experimenter tells the computer system to advance to the next point, registering simultaneously with a cross the place that corresponded to the fixation point. This process is repeated until the calibration output (cross marks) reproduces the 3 vertices of the triangle. At this moment, a validation procedure is made which consists in repeating the calibration procedure to be certain that the subject looked indeed at the 3 points. If not good, the validation should also be repeated. During all this process, if subject moves its head out from the chin-rest the process should be restarted.

After validation, the system informs the difference between the output of calibration and the output of validation in degrees of visual angles. If the difference is smaller than 5° degrees, the calibration is accepted and the experimental trials can start. Otherwise the calibration should be repeated. After accepting the calibration, the experimenter commands the system to start the trial.

Stimuli presentation

Every trial started with the presentation of a point of light in the middle of the screen (preceding the stimulus presentation) to drag the subject's gaze to the middle of the display area so in the beginning of every trial, all subjects were looking at the same place (the trial only starts if the system detects a fixation near this light point). Then the video starts and the system records the right eye movement of the subject during the 18 sec of the video. When the trial ends, another one starts or in case of the dog not being able to remain longer in the chin-rest, the dog is rewarded with food, and a new calibration is done before the next trial. Every time the subject moved its head out from the chin-rest (for example, when rewarded) a new calibration was performed until the 4 trials were completed. The data from each experiment were automatically stored in .edf format (EyeLink Data File).

2.1.7. Data preparation

To pull out the useful data from the recorded .edf files we used the MATLAB software (R2013a 8.1.0.604) from MathWorks, Inc. First, the MATLAB script converted the .edf files in ASC format, a text file format to make the files more accessible to the scripts in further processing steps. Following several scripts implementing the eye movement event detection algorithm developed by Lund Humanities Lab classified raw eye movement data consisted of X coordinate, Y coordinate and pupil diameter at a given time point in millisecond into eye

movement events such as blinks, saccade, glissade and fixation (Nyström & Holmqvist, 2010). The algorithm can identify eye movement events with little influence of human judgment as the algorithm automatically sets a noise-dependent velocity threshold that is adaptive to different individuals, different recording situations, and/or varying levels of noise. As a last step, the experimenter identified the area where each fixation resided. Duration of fixations output was then combined with the AOIs to produce a dataset of the total duration of fixations on different AOIs during the Choice Phase.

The Choice Phase lasted 8000ms but for some trials only a fraction of this time was recorded due to tracker loss (e.g. subject removed the head from the chin-rest). Since a partial recording may not be representative of a complete trial, trials with less than 1340ms successful eye-tracking during the Choice Phase were excluded (following Fletcher-Watson, Findlay, Leekam, & Benson, 2008). In addition, due to some experimenter errors, loss of data by the eye-tracker system, or incorrect eye-tracker calibration, further 14 trials were excluded, resulting in the exclusion of 16 trials of a total of 72.

2.1.8. Data analysis

As a measure of looking time to be compared between congruent and incongruent trials we used the duration of fixations, more precisely 3 variables:

- 1) Total duration of all fixations on the entire stimulus area,
 - 2) Proportion of duration of fixations on the actor's body area, calculated by dividing the total duration of all fixations on the actor's body area with the total duration of all fixations on the entire stimulus area,
 - 3) Proportion of duration of fixation on the chosen pot area, calculated in a similar way.
- Durations of fixating on the 2 AOIs were converted in proportions in order to reduce the noise on the data created as subjects strongly varied in the amount of time spent watching the videos.

Figure 2 shows the delimitation of the areas of interest mentioned above.

Data were analyzed in R version 3.3.2 for Windows through RStudio version 0.99.879 (R Core Team, 2016). For each variable, we analyzed the two blocks of trials separately, comparing the congruent and incongruent trials within each block, using two-tailed Wilcoxon matched-pairs signed-ranks tests. Blocks were analyzed separately because repeated presentations likely lead to familiarization or habituation to the stimuli (Wang et al., 2004) and consequently previous trials could influence the following ones, as dogs could lose interest in the videos or create new expectations of the choice outcome.

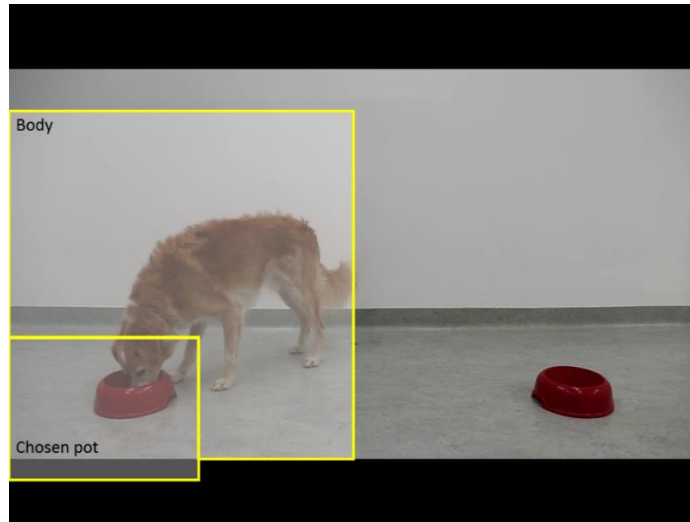


Figure 2. Chosen pot and body areas delimited by yellow lines.

2.2.Results

Total duration of all fixations on the video

For the first block of trials, we found no significant difference between the incongruent and congruent trials (Wilcoxon test: $N = 15$, $T(+) = 60.00$, $p > 0.9999$; Figure 3). Similarly, in the second block of trials, no significant difference between the incongruent and congruent trials was found (Wilcoxon test: $N = 9$, $T(+) = 24.00$, $p = 0.9102$; Figure 3).

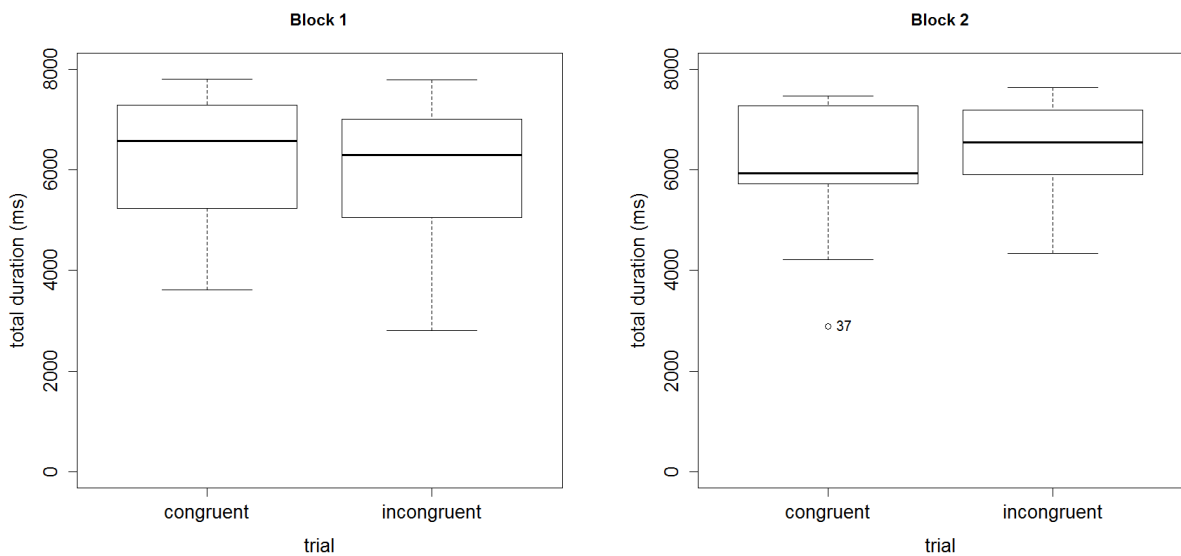


Figure 3. Total durations of all fixations on the video as function of congruency condition for Experiment 1. On the left, block 1 total durations; on the right, block 2 total durations.

Proportion of duration of fixations on actor's body area

Results of the analysis on proportion of duration of fixation on Michel's body area for block 1 revealed a statistically significant difference between congruent and incongruent trials ($N = 15$, $T(-) = -102.00$, $p = 0.015$), indicating that dogs looked longer in the incongruent trials than in the congruent ones. For block 2, we found no significant difference between the incongruent and congruent trials (Wilcoxon test: $N = 9$, $T(-) = -29.00$, $p = 0.4961$; Figure 4).

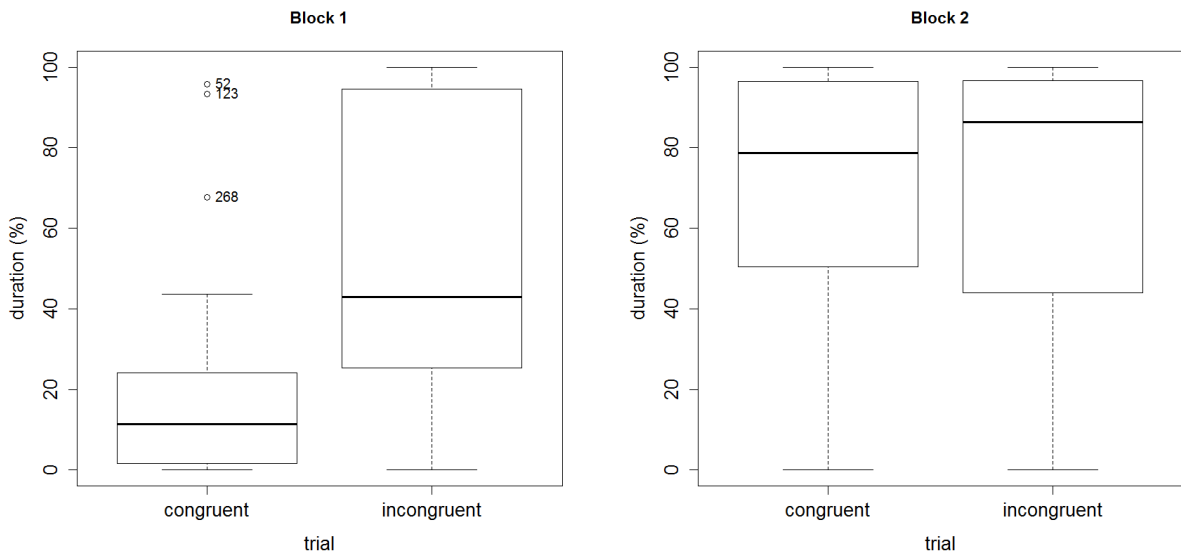


Figure 4. Proportion of durations of fixations on actor's body area as function of congruency condition for Experiment 1. On the left, block 1 proportions; on the right, block 2 proportions.

Proportion of duration of fixations on the chosen pot area

Results of the analysis on proportion of duration of fixation on chosen pot area for block 1 revealed a marginal difference between congruent and incongruent trials ($N = 15$, $T(+) = 93.00$, $p = 0.0637$; Figure 5), indicating a tendency for longer looking times in the congruent condition compared to the incongruent condition. For block 2, we found no significant difference between the incongruent and congruent trials (Wilcoxon test: $N = 9$, $T(+) = 23.00$, $p > 0.9999$; Figure 5).

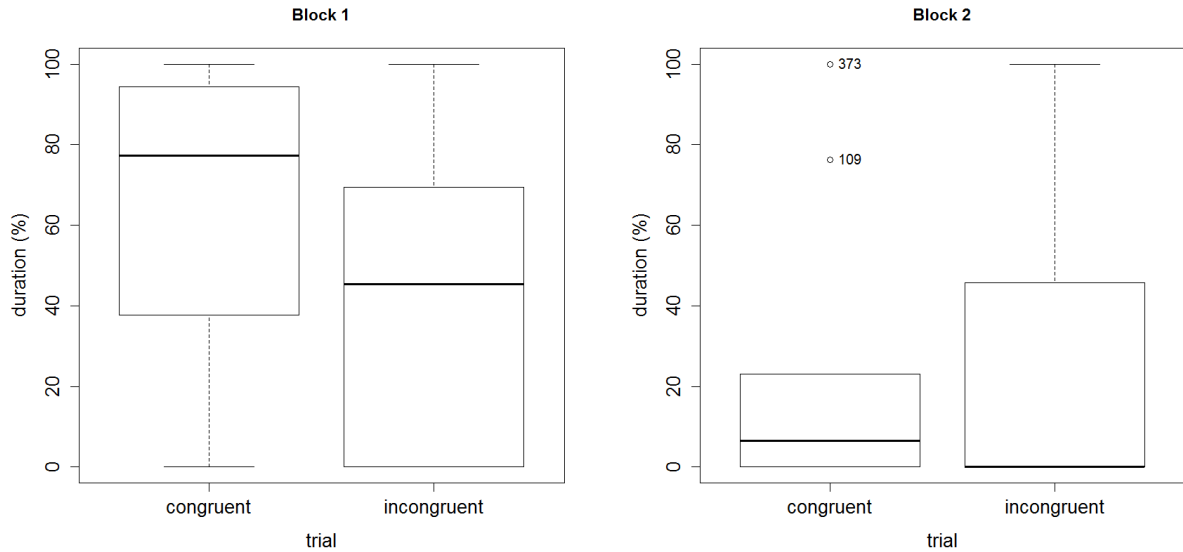


Figure 5. Proportion of durations of fixations on chosen pot area as function of congruency condition for Experiment 1. On the left, block 1 proportions; on the right, block 2 proportions.

2.3. Discussion

Although the subjects spent a similar amount of time watching the congruent and incongruent videos in both blocks, they showed different interest in different parts of the picture in the first congruent and incongruent test trials. In particular, we found that the dogs looked significantly longer at the area where Michel's body was standing in the incongruent trial compared to the congruent test trial, suggesting that seeing the actor dog eating from another pot than he had looked at beforehand was not expected. Interestingly, at the same time in the first block of test trials the dogs tended to show a greater interest in the chosen pot area in the congruent condition compared to the incongruent trial. That is, the dogs were more interested in a food location when Michel not only approached it but also looked at it beforehand. Taken together, these results suggests that the subjects not only formed a prediction about which pot Michel would approach based on his gaze direction in the gaze cue phase but they also seem to have recognized the referential content of his gaze (following Csibra, 2003; Csibra & Volein, 2008 - interpretation of experiments with human children).

Regarding the second block of trials, the negative results found could be due to a decrease in the sample size, from 15 to 9 subjects, that may not have allowed detecting the same effect of congruency as in block 1. Based on the graphs, however, it rather seems that the gazing pattern of the dogs indeed changed from block 1 to block 2. It seems that the subjects,

independently from condition, lost interest in the food locations and watched the actor dog longer. This might be due to the repeated presentation of the videos after which the subjects got no food reward from the pot presented. It is well-known that the outcome of the violation-of-expectation paradigm can strongly be influenced by repeated presentation of the same or similar stimuli which can influence the reaction to later presentations, as the subjects can lose interest in the videos or create new expectations of the outcome (Wang et al., 2004).

In sum, block 1 results seem to show evidence that dogs were able to predict Michel's next action based on his previous gaze direction. In order to investigate if dogs truly detect the link between Michel's gaze cue direction and his next action, that is, if the findings were indeed result of the congruence/incongruence between gaze and subsequent action, we run a second experiment where we compare the congruent and incongruent test trials with control trials where no pot directed gaze cue was presented in the initial part of the videos, but the subsequent action (eating from a pot) was still present.

3. Experiment 2

In the second experiment, our analysis strategy focused on the comparison of dogs' looking time between control and test trials. If dogs perceive directional gaze as a predictor of other's next action, then looking time results should be different between trials where a gaze cue is directed towards one of the pots (test trials), and the trials where no gaze cue is presented (control trials). Here, we hypothesize that if dogs perceive the connection between gaze direction and subsequent action, congruent trials should be the ones, between the three kinds of trial, with the smaller looking times, as they are the trials with the greater level of congruence between what happens in the first part of the videos and the final outcome. In control trials the outcome is also considered, from a human point of view, as unexpected since no directional predictor is presented by the actor dog and suddenly he appears in another place and in another position. In this line of thought, control trials should be the ones with longer looking times, given that in incongruent trials a potential predictor cue is present and can predict an action that is likely to occur, even if not in accordance with the indicated direction.

3.1. Methods

Methods were identical to Experiment 1, except the subjects and the stimuli presented that included control videos in addition to the ones from Experiment 1.

3.1.1. Subjects

Another eighteen dogs were recruited through the Clever Dog Lab dog database on a voluntary basis. Subjects were adult dogs, 10 females ranging from 1 to 10 years old ($M=5.40$, $SD=2.46$) and 8 males ranging from 3 to 6 years old ($M=4.50$, $SD=1.07$). All dogs had already participated in other eye-tracking studies (not in Experiment 1). Dogs were neither restrained nor forced to perform the tasks.

3.1.2. Stimuli

In addition to the test videos (congruent and incongruent videos) used in Experiment 1, control videos were also presented to subjects (Figure 6). Unlike the test videos, control videos did not have the gaze cue thus the actor dog remained looking forward through all the first part of the video until the black pause. The last part of the video clips was identical however: the actor dog ate either from the left pot (Control Left) or from the right pot (Control Right). This way, the control videos had 3 phases: No Cue Phase that lasted 8 secs, Black Pause Phase 2 secs and the Choice Phase 8 secs, all together making 18 secs for the total duration.

3.1.3. Experimental procedure

As in Experiment 1, two blocks of two trials were presented to subjects. The difference was that in each block, one of the videos was now a control video. Thus, one block contained a control video (Control Left or Control Right) and a congruent video (CL or CR), and the other block contained the second control video (Control Right or Control Left) and an incongruent video (IL or IR). The order of videos was counterbalanced between subjects, with the only rules that there were no repetition of the videos and no repetition of the choice side between the test videos. In this way, for example, if the first block had the Control Right and the CR videos, the second block would have the Control Left and the IL videos.

The procedure of stimuli presentation was identical to the Experiment 1.

3.1.4. Data preparation

Data preparation was identical to Experiment 1.

Regarding trials exclusion, 8 trials with less than 1340ms of data recorded were excluded. In addition, 4 trials were also excluded due to procedure errors, loss of data by the eye-tracker system, or incorrect eye-tracker calibration. In total, 12 trials were excluded from a total of 72 trials.

The measures of subjects' looking behavior used in the analyses were the same as in Experiment 1.

3.1.5. Data analysis

For each variable, we used 2 two-tailed Wilcoxon matched-pairs signed-ranks tests to compare control and congruent trials with the same choice outcome, and to compare the other control control and the incongruent trials, again with the same choice outcome.

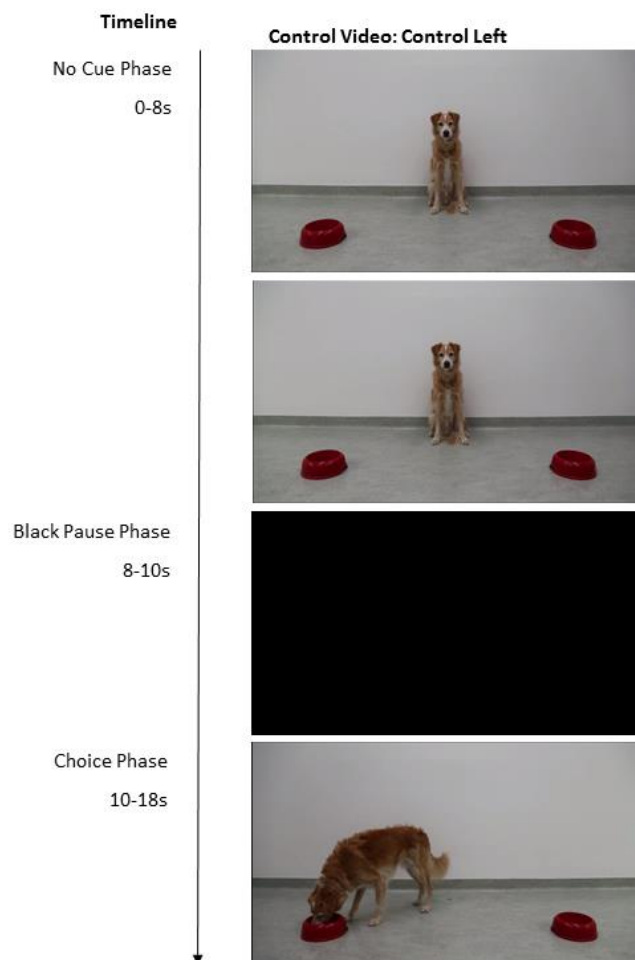


Figure 6: Control video timeline - screenshots from the 3 phases of the Control Left video: No Cue, Black Pause and Choice phase. Control Right video differs only in the Choice Phase, actor dog would be eating from the right side instead.

3.2. Results

Total duration of all fixations on the video

For control and congruent trials analysis on total duration of fixations, we found no significant difference between the two conditions (Wilcoxon test: $N = 13$, $T(+)$ = 67.00, $p > 0.1465$; Figure 7). For control and incongruent trials analysis, we also found no significant difference between the two conditions (Wilcoxon test: $N = 13$, $T(+)$ = 61.00, $p > 0.3054$; Figure 7).

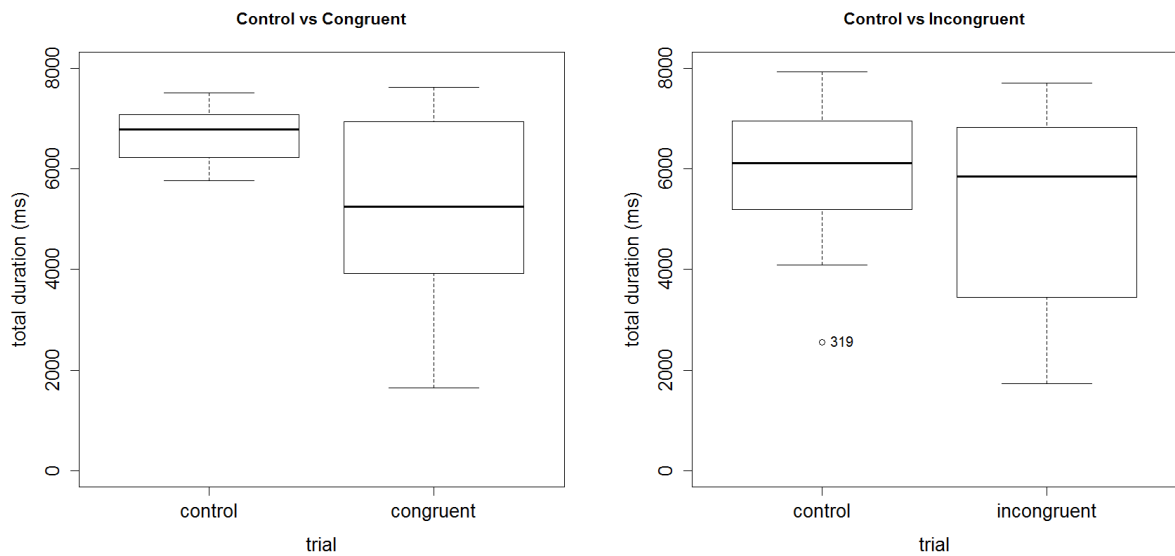


Figure 7. Total durations of all fixations on the video as function of test and control condition for Experiment 2. On the left, control and congruent trials; on the right, control and incongruent trials.

Proportion of duration of fixations on actor's body area

The analysis on proportion of looking time at Michel's body area showed no significant difference between control and congruent trials (Wilcoxon test: $N = 13$, $T(-)$ = -65.00, $p > 0.1909$; Figure 8). For control and incongruent trials, we found no difference between the two conditions (Wilcoxon test: $N = 13$, $T(+)$ = 53.00, $p > 0.6355$; Figure 8).

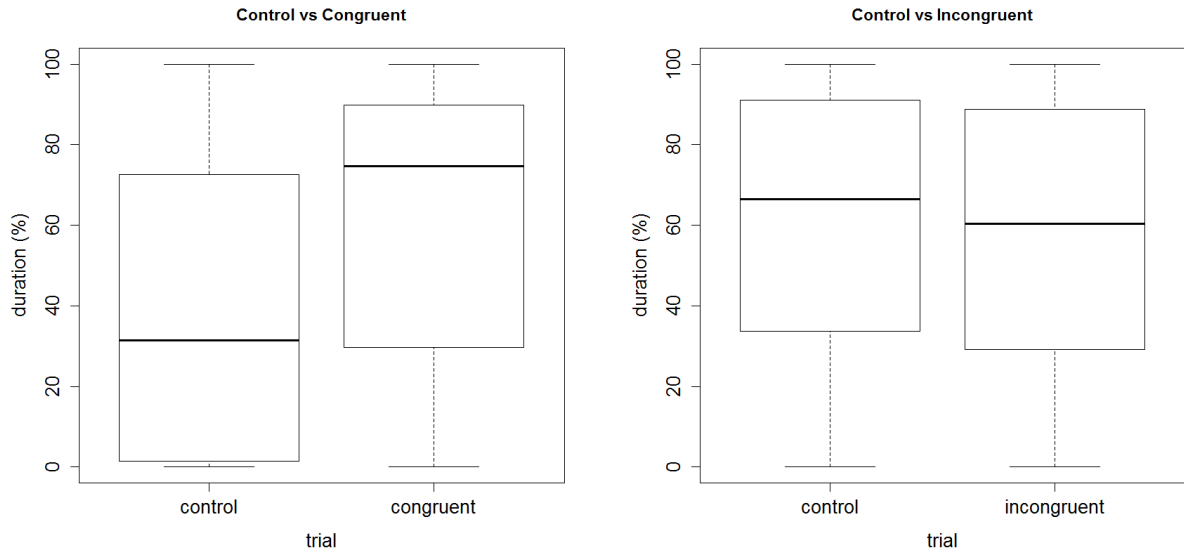


Figure 8. Proportion of durations of fixations on actor’s body area as function of test and control condition for Experiment 2. On the left, control and congruent trials; on the right, control and incongruent trials.

Proportion of duration of fixations on the chosen pot area

The analysis on proportion of duration of fixations on the chosen pot area for control and congruent trials revealed no significant difference between the two conditions (Wilcoxon test: $N = 13$, $T(+) = 54.00$, $p > 0.5879$; Figure 9). For control and incongruent trials, we found no significant difference between the two conditions (Wilcoxon test: $N = 13$, $T(-) = -52.00$, $p > 0.6848$; Figure 9).

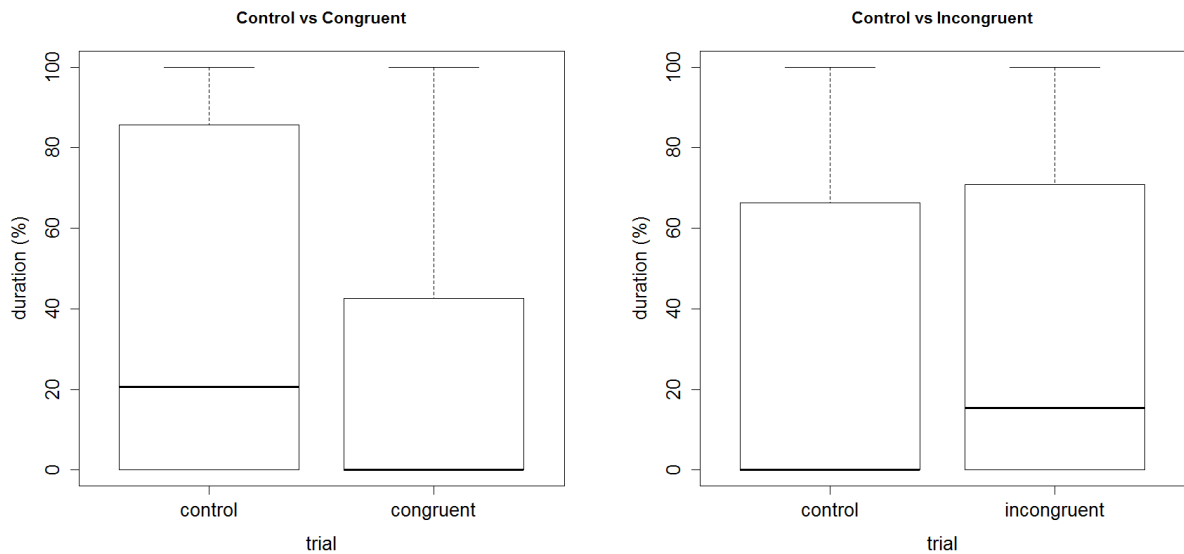


Figure 9. Proportion of durations of fixations on chosen pot area as function of test and control condition for Experiment 2. On the left, control and congruent trials; on the right, control and incongruent trials.

3.3. Discussion

In this experiment, we found a difference neither between the control and the congruent nor between the control and the incongruent trials. These results, in contrast to findings of Experiment 1, seem to suggest that dogs did not use the gaze cue to predict Michel's next action. It is possible however that other - methodological - factors explain the lack of significant differences in this experiment.

First of all, here we used data from 4 trials that might have already been too many presentations of similar videos. Experiment 1 indicated that the repeated presentation of the videos is likely to influence the looking times already from the third trial on, possibly eliminating the effect of congruency that was apparent in the first 2 trials. Dogs could have lost interest in the videos or created new expectations of the choice outcome, namely that it is as likely the actor choose the right or the left pot independently of which he looked at before. Moreover, we found high variations in the data that can be a result of the fact that being similar videos, the previous influenced the next ones. More detailed analyses that take the order of trials into account are needed.

In addition, the differences between control and test trials, especially for the control and incongruent analysis, are expected to be smaller than the ones between congruent and incongruent effect. Thus, if the effect is weak, our sample size might be too small to detect it. Improving sample size and using fewer number of trials for each subject can probably help to better answer our question addressed by Experiment 2.

4. General Discussion

This study was designed to test whether pet dogs are able to predict a conspecific's next action based on its gaze. Based on the findings of Experiment 1, dogs seem to have that capacity. Dogs looked reliably longer at the actor's body area when they previously saw the actor gazing towards another pot other than the pot that the actor approached afterwards, revealing that they had an expectation with which of the two pots the actor would be more likely to interact with.

In line with our findings, Duranton, Range, and Virányi (submitted) suggested that dogs interpret not conspecific but also human gaze in an object-choice situation as an intentional cue.

In that study, a human actor looked at the one of two containers (the baited one) and then subject dogs were allowed to choose a container. Interestingly, their results showed that dogs avoided the indicated container. Previous studies also found that dogs avoid a food location a human or a conspecific had looked at beforehand (Bálint et al., 2015; Soproni, Miklósi, Topál, & Csányi, 2001). Based on these findings, Duranton et al. (submitted) have proposed that dogs may interpret gaze as an intentional cue, or at least as a behavioural cue that reliably indicates others' further actions (*Intentional Gaze Cue Hypothesis*), that is, dogs may use this behavioural cue to predict that the next action of their partner looking at a food container will be to approach this location and to try to get the food. The results of the present study confirm this hypothesis.

This capacity develops in humans between 8 and 12-month-old infants (Phillips et al., 2002), and it is feasible to assume that other visually tuned social species also use gaze to predict others' next action. The ability to distinguish between physical and psychological objects and predict the latter's actions on other objects represents a highly adaptive capacity (Premack & Premack, 1997; Santos & Hauser, 1999). Surprisingly, however, a few studies found negative results were in squirrel monkeys and capuchin monkeys (Anderson et al., 2004). Based on our results we suggest that this was the case simply because these studies measured the overall attentiveness of the subjects. We also found that based on their total looking time subjects did differentiate between the congruent and incongruent conditions. However, looking at their gazing pattern in more details did reveal differences in their looking behaviour within the stimulus area. If former studies with negative results had used an eye-tracker and had been able to measure more precisely how long the subjects watched which part of the scene presented to them, possibly they might have found evidence that these primate species also create expectations about the others' future action based on their gaze (e.g. Anderson, Kuroshima, Kuwahata, & Fujita, 2004).

Moreover, our results revealed that dogs tend to pay more attention to the pot that the actor approached when this pot had previously been visually attended by the actor. It seems that the actor's gaze attracted the subjects' attention to the referred pot. Since the dogs differentiated between the congruent and incongruent trials into opposite direction in their fixations on the actor dog's body and the food location, it seems that they also recognized the referential content of Michel's gaze. This is in line with the results of a study by Soproni et al. (2001) where they investigated the responses of dogs to different types of directional human gestural cues in a procedure similar to that reported by Povinelli, Bierschwale, and Cech, (1999; Experiment 2). Dogs participated in an object-choice task where a human kneeling between two bowls, signaled the baited bowl in three different ways: turning her head and eye gaze toward the

correct bowl (at target trials), orienting her head and body in the same way as in the at target trials, but looking above the baited bowl to the upper corner of the room (above target trials) and orienting her head and body to the midline facing the dog and turning only her eye gaze toward the correct bowl (eyes only trials). In at target trials, body and gaze consistently signal the baited bowl, however above and eyes only trials have only one of this signal components consistently directed at the correct bowl. Above trials miss the referential component (gaze) and eyes only trials miss the body orientation component. Soproni et al. (2001) found that the dogs chose the correct bowl significantly above chance in the at target trials, but randomly on the above target and the eyes only trials, indicating that they took the referential component of human gaze (including face and body orientation) into account. Less relevant for our study is that they failed to choose solely based on the orientation of human eyes. This is actually similar to the performance of children and chimpanzees, as reported by Povinelli et al. (1999) who hypothesized that the direction of the human's eyes was too inconspicuous for the subjects.

Both of our findings (that dogs can predict others' action and may infer the presence of a referent based on their gaze) raise the question what kind of mechanisms underlie the dogs' performance. Do they interpret others' gaze in a mentalistic way or do they make direct inferences about their behaviour? Studies in pre-verbal infants have suggested different mechanisms to explain similar behavioural performance. For instance, a mentalistic interpretation (e.g. Luo & Baillargeon, 2010; Scott & Baillargeon, 2009) of infant gaze following behavior has argued: *“That person is looking at something in particular. Let me look too, so that I can see what she/he is seeing. Let me understand what this person is thinking and feeling about what she/he sees, so that I can predict their behavior”* (Goswami 2006).

Moll & Tomasello (2004) found that 12 and 18-months-old infants move their position to look behind a barrier if an adult is looking at something behind that barrier, out of the view of the infant. These results are often used to show that seeing is a mental act, since it appears that infants moved because they wanted to see what the other was seeing.

However, an alternative interpretation for these results, mentioned by Csibra & Volein (2008), is that infants approached the looked spot simply because the looking behavior of the adult had drawn their attention to that hidden part of the space, and their action provided no evidence if they inferred that an object should be found there. This “lean” interpretation argues that the movement of the adult's head draws the infant's attention to a section of space. Infants may coincidentally look at the same object as the adult because they notice the most salient object in their visual field to which they have been drawn. This would not need any understanding about the adult's attention or a reference to the object looked (Butterworth & Jarrett, 1991;

Moore, 1994). According to this interpretation, gaze following is a simple reflexive orientation behaviour (Shepherd, 2010). However, this does not mean that infants are not capable of drawing inferences about the goal-directed actions of others.

This study provides no evidence that dogs would infer their social partner's referential intention or intention to make a next action based on their gaze direction. Further research has to clarify whether dogs are capable of making inferences about others' intention. Studies so far found contradicting results. Marshall-Pescini, Ceretta, & Prato-Previde (2014) investigated whether dogs attribute intentions to an animate (a human) but not an inanimate (a black box) agent interacting with an object. Their results provided the first evidence that a non-primate species can perceive another individual's actions as goal-directed. Still, it remains an open question whether this interpretation relies on attributing intentions to humans. Arguing against it, Riedel, Buttelmann, Call, and Tomasello (2006) found that dogs did not differentiate intentional from accidental manner of placing a marker that identified the location of hidden food in an object-choice situation.

Strengths and limitations of the study

In comparison to other studies using a similar violation-of-expectation paradigm (Phillips et al., 2002; Santos & Hauser, 1999), although we found some effects of congruency on the dogs' gazing behaviour these effects were relatively weak. This is perhaps because our study misses the familiarization or habituation phase. In the familiarization phase subjects are familiarized with the situation subsequently presented in the test phase, generally, differing from the test trials in the direction of the gaze cue or in the object where actors' act on. However, we intended to investigate if subjects already had the (studied) expectation before participating in the experiment, and familiarization trials rule out this possibility. This weak effect may also be due to the fact that in our study the dogs were kind of "restrained" to a chin rest position which was necessary for eye-tracking but might have prevented the dogs from moving their heads freely and thus from looking away from the stimulus more easily.

Still, using an eye-tracker was a great advantage of our study, as it allowed a much more detailed and exact analysis of how long and where subjects looked within the stimulus area. This actually allowed for detecting that dogs did differentiate between the congruent and incongruent conditions, as the overall looking time of dogs was similar across conditions.

5. References

- Anderson, J., Kuroshima, H., Kuwahata, H., & Fujita, K. (2004). Do squirrel monkeys (*Saimiri sciureus*) and capuchin monkeys (*Cebus apella*) predict that looking leads to touching? *Animal Cognition*, 7(3), 185–192. <http://doi.org/10.1007/s10071-004-0209-z>
- Bálint, A., Faragó, T., Meike, Z., Lenkei, R., Miklósi, Á., & Pongrácz, P. (2015). “Do not choose as I do!” – Dogs avoid the food that is indicated by another dog’s gaze in a two-object choice task. *Applied Animal Behaviour Science*, 170, 44–53. <http://doi.org/10.1016/j.applanim.2015.06.005>
- Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology*, 38(6), 958–966. <http://doi.org/10.1037//0012-1649.38.6.958>
- Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9, 55–72. <http://doi.org/10.1111/j.2044-835X.1991.tb00862.x>
- Call, J., Bräuer, J., Kaminski, J., & Tomasello, M. (2003). Domestic dogs (*Canis familiaris*) are sensitive to the attentional state of humans. *Journal of Comparative Psychology*, 117(3), 257–263. <http://doi.org/10.1037/0735-7036.117.3.257>
- Cooper, J. J., Ashton, C., Bishop, S., West, R., Mills, D. S., & Young, R. J. (2003). Clever hounds: social cognition in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, 81(3), 229–244. [http://doi.org/10.1016/S0168-1591\(02\)00284-8](http://doi.org/10.1016/S0168-1591(02)00284-8)
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1431), 447–58. <http://doi.org/10.1098/rstb.2002.1235>
- Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26(1), 1–11. <http://doi.org/10.1348/026151007X185987>
- Duranton, C., Range, F., & Virányi, Z. (submitted). Do pet dogs (*Canis familiaris*) follow ostensive and non-ostensive human gaze to distant space and to objects?

- Fitch, T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language : constructing cognitive phylogenies. *Neuron*, 1–21.
<http://doi.org/10.1016/j.neuron.2010.03.011>
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37(4), 571–583.
<http://doi.org/10.1068/p5705>
- Gácsi, M., Miklósi, Á., Varga, O., Topál, J., & Csányi, V. (2004). Are readers of our face readers of our minds? Dogs (*Canis familiaris*) show situation-dependent recognition of human's attention. *Animal Cognition*, 7(3), 144–153. <http://doi.org/10.1007/s10071-003-0205-8>
- Goldman, A. I. (2012). Theory of Mind. In E. Margolis, R. Samuels, & S. Stich (Eds.), *Oxford Handbook of Philosophy and Cognitive Science*.
- Goswami, U. (2006). The foundations of psychological understanding. *Developmental Science*, 9(6), 545–50. <http://doi.org/10.1111/j.1467-7687.2006.00531.x>
- Hare, B., & Tomasello, M. (1999). Domestic dogs (*Canis familiaris*) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology*, 113(2), 173–177. <http://doi.org/10.1037//0735-7036.113.2.173>
- Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, 9(9), 439–44. <http://doi.org/10.1016/j.tics.2005.07.003>
- Herendy, V. (2011). *With whom and how? Investigating the mechanisms adult dogs use to locate hidden food based on conspecific and human gaze as well as human pointing in an object choice task*. Internship's report at Clever Dog Lab, Vienna.
- Horowitz, A. (2009). Attention to attention in domestic dog (*Canis familiaris*) dyadic play. *Animal Cognition*, 12, 107–118.
- Luo, Y., & Baillargeon, R. (2010). Toward a mentalistic account of early psychological reasoning. *Current Directions in Psychological Science*, 19(5), 301–307.
<http://doi.org/10.1177/0963721410386679>
- Marshall-Pescini, S., Ceretta, M., & Prato-Previde, E. (2014). Do domestic dogs understand human actions as goal-directed? *PLoS ONE*, 9(9), e106530.
<http://doi.org/10.1371/journal.pone.0106530>

- Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Z., & Csányi, V. (2003). A Simple reason for a big difference: wolves do not look back at humans, but dogs do. *Current Biology*, *13*, 763–766. <http://doi.org/10.1016/S>
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*(1), F1-9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15323111>
- Moore, C. (1994). Social Understanding at the End of the First Year of Life. *Developmental Review*, *14*(4), 349–372. <http://doi.org/10.1006/drev.1994.1014>
- Müller, C. a, Mayer, C., Dörrenberg, S., Huber, L., & Range, F. (2011). Female but not male dogs respond to a size constancy violation. *Biology Letters*, *7*(5), 689–91. <http://doi.org/10.1098/rsbl.2011.0287>
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204. <http://doi.org/10.3758/BRM.42.1.188>
- Péter, A., Miklósi, Á., & Pongrácz, P. (2013). Domestic dogs' (*Canis familiaris*) understanding of projected video images of a human demonstrator in an object-choice task. *Ethology*, *119*(10), 898–906. <http://doi.org/10.1111/eth.12131>
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78. [http://doi.org/10.1016/S0010-0277\(02\)00073-2](http://doi.org/10.1016/S0010-0277(02)00073-2)
- Povinelli, D. J., Bierschwale, D. T., & Cech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology*, *17*(1), 37–60. <http://doi.org/10.1348/026151099165140>
- Premack, D., & Premack, A. J. (1997). Infants attribute value \pm to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, *9*(6), 848–856. <http://doi.org/10.1162/jocn.1997.9.6.848>
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Riedel, J., Buttelmann, D., Call, J., & Tomasello, M. (2006). Domestic dogs (*Canis familiaris*) use a physical marker to locate hidden food. *Animal Cognition*, *9*(1), 27–35. <http://doi.org/10.1007/s10071-005-0256-0>

- Santos, L. R., & Hauser, M. D. (1999). How monkeys see the eyes: Cotton-top tamarins' reaction to changes in visual attention and action. *Animal Cognition*, 2(3), 131–139. <http://doi.org/10.1007/s100710050033>
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80(4), 1172–1196. <http://doi.org/10.1111/j.1467-8624.2009.01324.x>
- Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, 4(March), 5. <http://doi.org/10.3389/fnint.2010.00005>
- Somppi, S., Törnqvist, H., Hänninen, L., Krause, C., & Vainio, O. (2012). Dogs do look at images: eye tracking in canine cognition research. *Animal Cognition*, 15(2), 163–74. <http://doi.org/10.1007/s10071-011-0442-1>
- Soproni, K., Miklósi, A., Topál, J., & Csányi, V. (2001). Comprehension of human communicative signs in pet dogs (*Canis familiaris*). *Journal of Comparative Psychology*, 115(2), 122–126. <http://doi.org/10.1037/0735-7036.115.2.122>
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–735. <http://doi.org/10.1017/S0140525X05000129>
- Wang, S.-H., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: evidence from violation-of-expectation tasks with test trials only. *Cognition*, 93(3), 167–98. <http://doi.org/10.1016/j.cognition.2003.09.012>
- Werhahn, G., Virányi, Z., Barrera, G., Sommese, A., & Range, F. (2016). Wolves (*Canis lupus*) and Dogs (*Canis familiaris*) differ in following human gaze into distant space but respond similar to their packmates' gaze. *Journal of Comparative Psychology*, 130(3), 288–298.
- West, R. E., & Young, R. J. (2002). Do domestic dogs show any evidence of being able to count? *Animal Cognition*, 5(3), 183–6. <http://doi.org/10.1007/s10071-002-0140-0>
- Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6(3), 297–311. <http://doi.org/10.1111/1467-7687.00286>

ANNEXES
Annex I – Experimental set-up

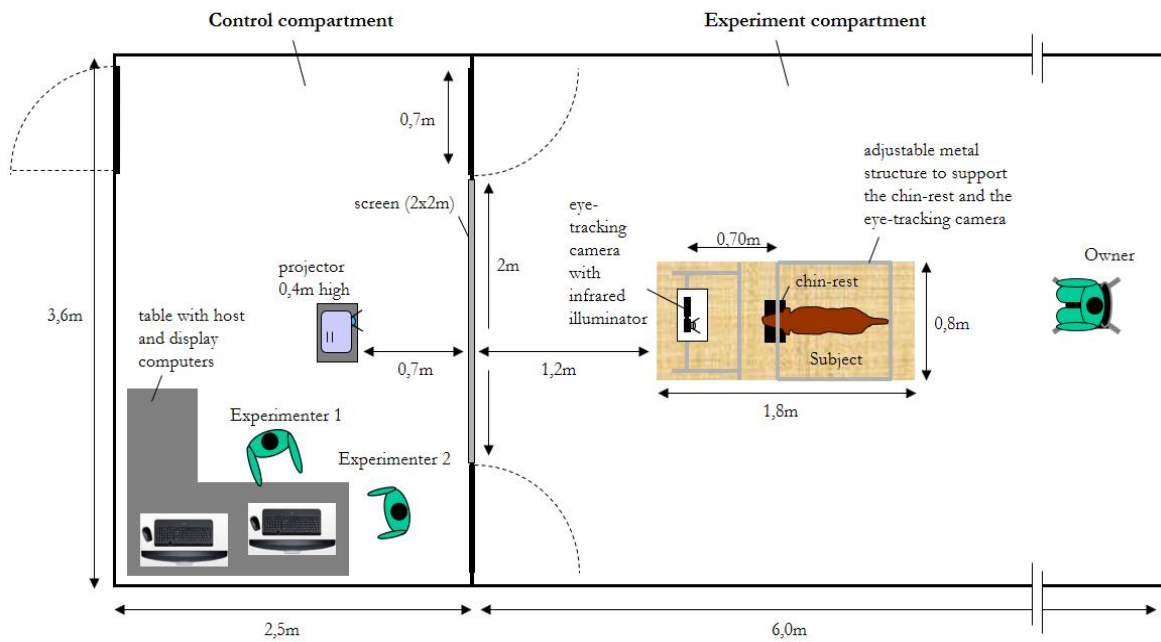


Figure A1. Top view of the experimental room

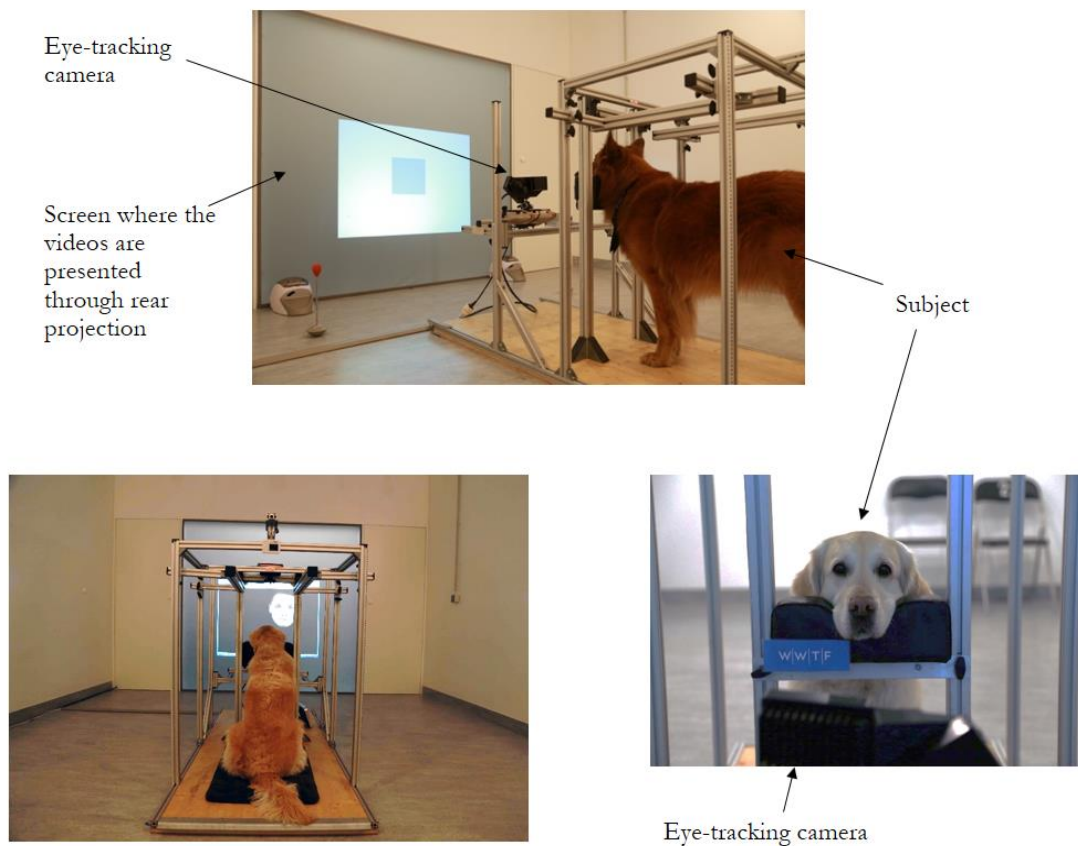


Figure A1: General view of the experiment compartment with subjects in the chin-rest