Evaluating the potential of artificial visual stimulus presentation in object search tasks for dogs

András Péter

Eötvös Loránd University, Graduate School of Biology
Head: Dr Anna Erdei, DSc

Ethology Graduate Program
Head: Dr Ádám Miklósi, DSc

Supervisor:
Dr Péter Pongrác, Habil.
Associate professor

Eötvös Loránd University, Department of Ethology

2016
# TABLE OF CONTENTS

## INTRODUCTION

1. **Artificial Visual Stimuli** 1
   
   1.1. Artificial Visual Stimuli in Behavioral Experiments 1
   
   1.2. Advantages of Utilizing Images as Visual Stimuli 1
   
   1.3. Disadvantages of Utilizing Images as Visual Stimuli 3
   
   1.4. Image Processing Modes and Referential Understanding 6
   
   1.5. Testing Picture-Object Recognition in Animals 8

## Dogs (*Canis familiaris*) 10

2.1. Studies of Companion Dog Cognition in the Last Two Decades 10

2.2. Studies of Dog Cognition Utilizing Images 12

2.3. The Visual Perception of Dogs 15

2.4. What Can Be Inferred About Dogs’ Image Processing From Studies to Date 19

2.5. Studies Focused on the Referential Abilities of Dogs 21

## STUDIES 23

**Study 1: The Effect of Ostensive Cues on Perseverative Errors in a Visible Displacement Task** 26

Aims and Questions 26

Methods 29

Results 34

Summary 37

**Study 2: Image Processing in a Spontaneous Information Transfer Task** 39

Aims and Questions 39
# TABLE OF CONTENTS

- **METHODS**  
  - page 40
- **RESULTS**  
  - page 47
- **SUMMARY**  
  - page 49
- **STUDY 3: IMAGE PROCESSING IN A VIOLATION OF EXPECTATION TASK**  
  - page 52
  - **AIMS AND QUESTIONS**  
    - page 52
  - **METHODS**  
    - page 55
  - **RESULTS**  
    - page 60
  - **SUMMARY**  
    - page 67
- **STUDY 4: VIRTUAL VISIBLE DISPLACEMENT TASK ON THE TOUCH-SCREEN**  
  - page 70
  - **AIMS AND QUESTIONS**  
    - page 70
  - **METHODS**  
    - page 73
  - **RESULTS**  
    - page 82
  - **SUMMARY**  
    - page 85
- **GENERAL DISCUSSION**  
  - page 88
- **SUMMARY**  
  - page 96
- **ACKNOWLEDGEMENTS**  
  - page 98
- **REFERENCES**  
  - page 99
- **ABSTRACT**  
  - page 116
- **ÖSSZEFoglaló**  
  - page 117
- **DOI ADATLAP**  
  - page 118
INTRODUCTION

1. Artificial visual stimuli

1.1. Artificial visual stimuli in behavioral experiments

From the dawn of animal behavior studies, artificial visual stimuli were an indispensable instrument in the scientific toolbox of researchers. In 1936 Noble and Clausen studied the aggregation behavior of De Kay's snakes (*Storeria dekayi*) by using colored wax casts of the species, whereas Tinbergen and Perdeck (1950) used dummies for the investigation of the releasing stimuli of herring gull chicks’ (*Larus argentatus*) begging behavior as did Rilling and colleagues (1959) to study prey recognition in praying mantis (*Parastagmatoptera unipunctata*).

A popular form of artificial visual stimuli are images. Images can be either static (e.g. photos, drawings) or dynamic (e.g. videos, animations). Working with images has gained traction in almost every field of animal behavior studies. Social behavior of hens (*Gallus gallus domesticus*) was studied with the help of photographs of conspecifics (Dawkins 1996), fear reaction of ewes (*Ovis aries*) was analyzed with life-sized projected photographs of humans (Vandenheede and Bouissou 1994). Videos were also used to address a wide range of topics e.g.: species recognition in anolis lizards (*Anolis graham*; Macedonia and Stamps 2010), interspecies signaling between humans and dogs (Pongrácz et al. 2003) and mate choice of the amazon molly (*Poecilia formosa*; Landmann et al. 1999). Computer animations were successfully employed among other things for investigating the prey catching behavior of green frogs (*Rana clamitans*; Roster et al. 1995) and for the analysis of tail-flick signals in Jacky dragons (*Amphibolurus muricatus*; Peters and Evans 2003).

Utilizing images is also an almost mandatory element of the investigation of the mechanism and limits of the visual processing of animals (for a review see Bovet and Vauclair 2000). The popularity of using images as artificial stimuli can be attributed to the numerous benefits of this method compared to using natural or other types of artificial visual stimuli.

1.2. Advantages of utilizing images as visual stimuli

The advantages of images as artificial stimuli can be summarized in three main points: repeatability, stimulus control and ease of use. Repeatability means that by using images the
INTRODUCTION

exact same stimulus can be presented to all subjects. This uniformity encompasses spatial configuration, visual detail and in case of dynamic images also movement patterns and timing. By using images in studies involving social interactions it is possible to present a social partner with behavior unaffected by the subject (Macedonia et al. 1994). This invariability provided by images is a large step towards eliminating the Clever Hans effect (Pfungst 1911), which is the involuntary cueing of the subjects, that biases their behavior.

Using images in an experiment enables the manipulation of the presented stimuli which is a key element of scientific studies. These manipulations can entail the modification of color (Kodric-Brown and Nicoletto 2001), pattern, shape or the reduction of a more complex image to a few key features. In case of videos even the behavior sequence (Clark and Uetz 1992) or speed of movement (Rowland 1995) can be altered. By using computer generated images it is possible to overcome the reliance on pre-recorded material, and any novel stimulus can be created with all the visual parameters controlled by the experimenters. For example Woo (2007) created a 3D computer animated model of a Jacky dragon (*Amphibolurus muricatus*) and by making it perform entirely new motion sequences could successfully study the social behavior of the species.

The manipulation if images makes it possible to alter certain elements of the stimulus without affecting others. Using video sequences of the courtship behavior McKinnon (1995) manipulated the throat color of male three-spined sticklebacks (*Gasterosteus aculeatus*) when studying mate preference of the females. Another compelling example are studies with point-light displays where the visual stimuli are solely comprised of moving dots (Johansson 1973). These dots are arranged in a configuration that roughly corresponds to an animal (e.g. human, pigeon or dog) and move in a way that follows the movement of the model animal. This arrangement enables the study of the perception of biological motion without the confounding effects of color, pattern or shape recognition. One of the key advantage of using images is the ability to present a stimulus that would not occur in nature and would be difficult to produce otherwise in reality (e.g. presenting impossible events in a violation of expectation test).

Other artificial stimuli (e.g. dummies or models) are also characterized by the advantages of repeatability and stimulus control, however from the aspect of ease of use (production and presentation) images are superior. One should only consider how easy it is to take (digital) pictures or videos and present them via a large number of mediums (e.g. prints,
projections, monitors). Video playbacks enable to present precisely timed complex temporal patterns that would otherwise be impossible with static models or would require the construction of complicated machinery (e.g. Stout and Amlaner 1978). The modification of images can be achieved with commercially available software and does not require specific training, whereas creating dummies that replicate sufficiently well their original template requires specialized skills of craftsmanship.

1.3. Disadvantages of utilizing images as visual stimuli

Despite the numerous benefits described in the previous section, it is necessary to consider certain implications that arise from the use of images. Albeit it was mentioned earlier among their advantages, the non-interactive nature of images can be regarded as a drawback as well. If the nature of the studied situation requires that the presented stimulus responds to the subject (e.g. studying social interactions), it can lead to abnormally heightened (e.g. opponent not yielding in an aggressive interaction) or lowered (habituating to the unmoving or repeating stimulus) reaction (Patterson-Kane et al. 1997). It is possible to overcome this limitation by allowing the presented stimulus to be interactive and to change in response to the behavior of the subject. This can be achieved by having a pool of pre-recorded videos and always displaying the adequate one depending on the subjects reactions or to have a computer generated stimulus being created in real time (Woo 2007).

Another issue that affects pre-recorded images is pseudoreplication (Rosenthal 1999). If all subjects are presented with the same stimulus it is possible that they react to a detail of the images that is specific to that one image only. Such adverse effects can be mitigated by carefully screening the visual stimuli for unbalanced visual elements and removing all unnecessary components. Another method is to use computer generated images, which would allow perfect control over all relevant details. Pseudoreplication can be also eliminated by including a certain level of variability in the generation of the stimuli, so that every subject is faced with a slightly different stimulus.

The most obvious difference to other types of stimuli is that images are inherently two dimensional and therefore lack certain visual features that contribute to depth perception. Depth

---

1 The recent advancement of commercially available 3D printers might make the production of models as simple as creating and presenting images, however as of 2016 objects produce by 3D printers are static and have only one or few colors.
perception can be guided by many types of visual cues, some cues are monocular (can be detected by using one eye) and some are binocular (only available if some part of the animal’s visual field is perceived by both eyes). All types of binocular depth cues are absent from images: stereopsis (calculating an object’s distance from the differences of the two retinal images), convergence (judging the distance of an object by detecting the convergence of extraocular muscles when the eyes are directed towards the target) and shadow stereopsis (retinal images with different shadows) can impart depth perception (Puerta 1989).

From monocular depth cues the following would be unavailable in images: accommodation (detecting the state of the ciliary muscles in the eye that stretch the lens while focusing on an object) and motion parallax (a moving observer can judge the relative distance of stationary objects from their apparent movement against a background). Other monocular depth cues can be reproduced by two dimensional stimuli: occlusion (an object that is closer could partially block the view of an object that is further away), relative size (from two identical object with a known size the one having a smaller retinal image will be perceived as being further away), familiar size (the brain can calculate the distance of an object with a known size from the area it’s retinal image occupies), perspective (parallel lines seem to converge in the distance which can contribute to a sense of depth), texture gradient (on nearby objects fine details are more discernible than on distant ones), lights and shadows (the way light is reflected by objects and the shadows cast by them provides cues about their shape and spatial position) and depth from motion (a moving object that becomes smaller on the retina will be perceived as moving away, and vice versa). An additional problem of using 2D stimuli is that as the animal moves relative to the image the shape of objects on it will be distorted. However this can be prevented by choosing an appropriate experimental design (e.g. limiting the distance the animal can approach the screen).

As it can be seen from the previous enumeration of depth cues it is obvious that by viewing an image an animal would perceive some of these cues correctly but others as missing or conflicting. Many animals have the potential to perceive depth cues but they do not always use these (e.g. chicken; Evans et al. 1993). However if depth cues (especially binocular) are an essential part of the question under investigation the use if images is not a suitable option to conduct the experiments. Luckily for the researchers, still a handful of depth cues remain available in the 2D images, and in many stimulus presentation scenarios depth cues do not play a crucial role.
Another critical point that has to be considered when using images is that the commercially available picture-recording and reproducing devices have been constructed to match with the capacity and limitations human visual system (Fleishman et al. 1998). Therefore animals whose visual system has differing parameters from that of humans’ could perceive images differing from reality. Color reproduction is one of the areas where special care has to be taken when planning to utilize images.

The human visual system relies on three different photoreceptor cells (cones) to distinguish colors (humans are trichromats). Each photoreceptor type contain photopigments that absorb lights at different wavelengths. Depending on the species, animals can be monochromats with no color perception (e.g. nocturnal monkeys, *Simia trivirgata*; Jacobs et al. 1993b), dichromats with being able to distinguish less colors than humans (e.g. dogs, *Canis familiaris*; Neitz et al. 1989) and tetrachromats that can distinguish more colors than humans (e.g. pigeons, *Columba livia*; Emmerton and Delhis 1980). Independently from the number of different photoreceptive cell types an animal could also be able to detect wavelengths of light not visible to humans. Rats (*Rattus norvegicus*) for example are dichromats but can detect ultra violet light (Jacobs et al. 2001).

Another attribute of the visual system to consider when presenting videos is the critical flicker-fusion frequency (CFF). CFF is the frequency at which a flickering image starts to appear continuous to an observer. The maximum CFF of humans is between 50 and 60Hz under optimal conditions (Landis 1954), however it can become lower than 10Hz in low light conditions (Hecht 1933). Diurnal and fast moving animals have typically high CFF as the tsetse fly (*Glossina morsitans*) with 200Hz (Miall 1978), while nocturnal species or animals living in low light environments have low CFF values, as the nocturnal gecko (*Hemidactylus turcicus*) with 18Hz (Dodt 1961). Visual acuity is another factor that could affect perception of images. Animals with high visual acuity might not perceive the scene or object depicted by an image but only the building blocks of it (pixels, pigment dots, etc.). This is especially true if the animal is allowed to get close to the image or screen.

Considering all the previously listed perceptual attributes it is clear that it is crucial to precisely know the parameters of visual perception of the target species when planning to employ images in an experiment. If the animal’s visual perception is in the range of human perception (e.g. cannot distinguish more colors, does not have a higher CFF) than commercially available tools could suffice to create adequate images. However if the animal has perceptual abilities that exceeds that of humans then special equipment might be required. In case the
animal having a higher CFF special high speed cameras and displays could be used, or in case of higher visual acuity high resolution devices would be necessary. Luckily such equipment is relatively easy to obtain nowadays compared to even a decade earlier.

Still it could be impossible to display images in a way that would present an adequate reproduction of reality to the animal’s visual system (e.g. for animals that can perceive UV light). There are however a number of studies where animals responded appropriately to stimuli that were not perfectly matching their visual system (Clark and Uetz 1990; Macedonia et al. 1994). This might have been because other cues (shape, movement, brightness differences, etc.) could have provided sufficient information for recognition. One has only to remember that humans can recognize black and white pictures or videos that lack color information. Therefore it is clear that being able to recognize the content of an image does not solely depend on the capacity of the visual system but also on the way the animal cognitively processes artificial visual stimuli (Fagot and Parron 2010). This question in turn leads to the topic of picture-object recognition and image processing modes that will be discussed in the next chapter.

1.4. Image processing modes and referential understanding

According to Fagot and colleagues (2010) animals can process images in three different ways: independence mode, confusion mode and equivalence mode. Processing modes describe the cognitive connection the animal makes between the image and the object or scene it depicts. In independence mode the animal makes no connection between the picture and its content, but processes the picture as a combination of features or patterns independently of what the picture might represent. In confusion mode the animal confuses the image and its referent, thus reacting the same way to an image as to the real object. In equivalence mode the animal comprehends that the picture is a representation of the depicted object. This latter level corresponds to referential understanding, which is the ability to perceive an object (picture, video, replica, scale model, action, etc.) as standing for another entity in the world (DeLoache 1991; Gliga and Csibra 2009).

Mirror self-recognition experiments provide a good illustration of how the three levels of image processing manifests (Gallup 1968). When an animal is presented with a mirror it can show three types of responses: a) do not react to the mirror in any species specific way, meaning that it did not perceive its mirror image as either a conspecific or itself (independence mode); b) it can react as if seeing an unknown conspecific (e.g. threatening it) which would correspond
to confusion mode; c) finally it can show self-directed behavior which is an indication that it recognized itself in the mirror (equivalence mode).

Whether it is necessary to establish if an animal can recognize the connection between images and their real world referent when using artificial stimuli is a debated question. According to Soto and Wasserman (2010) when the researchers’ aim is to study animal vision and visual processing, it is sufficient to show that the animal is able to solve a given visual task (e.g. identifying or categorizing an image). According to their reasoning, because animals are constantly faced with similar visual challenges in their natural environment the results derived from laboratory experiments can show how biological systems adapted to their environment process visual information.

According to opponents of this view it is important to know whether an animal perceives the correspondence between the information in images and the information in the real world. They argue that only by knowing this are researchers able to determine what kind of recognition and categorization processes are tested when using artificial visual stimuli (Spetch 2010). For example if in a categorization study the animals perceive images in confusion or equivalence mode then results would tell us about how they categorize the real objects that were on the images. On the other hand if the animals processed the images in independence mode than the results would only let us know how they categorize unfamiliar visual stimuli that have a particular color, structure and shape.

Both views have their merits considering to the aim of a given study. As pointed out by Lazareva (2010) in studies of basic mechanisms of learning, categorization, and vision it is not necessary to ascertain whether the animal connects the images with reality (additionally in a number of cases such experiments rely on stimuli that do not resemble real world object in which case the question of correspondence is rendered mute). Whereas when images are used to study the formation of ‘natural categories’ or as direct replacement of natural stimuli then it is crucial to establish whether animals recognize the objects on the images. There are a number of different ways how processing of images can be tested in animals, six of the most widely employed methods will be introduced in the next chapter.
1.5. Testing picture-object recognition in animals

Studies of how animals relate images to the objects depicted on them are commonly referred to as picture-object recognition tasks. They can be divided into two main groups: a) some studies investigated the subjects’ spontaneous responses towards artificial stimuli, b) whereas another group of methods is based on acquired responses: in these the animal is always trained first to discriminate or categorize a certain set of stimuli.

One method that relies on spontaneous responses aims to observe the occurrence of natural behavior when the animal is presented with artificial stimuli. If the animal shows similar behavior to the artificial stimulus as to its real counterpart, then it is assumed that it perceived the stimulus as being ‘the same’ as the real one (confusion mode). This method was successfully used with e.g.: rhesus monkeys (Macaca mulatta; Sackett 1965), dogs (Canis familiaris; Fox, 1971), jumping spiders (Maevia inclemens; Clark and Uetz 1990) or ewes (Ovis aries; Bouissou et al. 1996).

Another similar method is measuring the preference shown for one of two visual stimuli. This method was employed with e.g.: pig-tailed macaques (Macaca nemestrina; Fujita 1993) and green swordtail fish (Xiphophorus helleri; Rosenthal et al. 1996). If the animals show similar preference patterns for real and for artificial stimuli pairs, then one can conclude that the animal perceived the artificial stimuli as being ‘the same’ as the real ones (confusion mode). The interpretation of both of the previously introduced methods can be challenged that a ‘natural reaction’ or preference could also be triggered by some key perceptual element of the stimuli, without an actual recognition of the content of the images (independence mode).

One type of the studies that rely on acquired responses compares the speed with which an animal learns to assign objects to natural or artificial categories (e.g. pigeons, Columba livia; Herrnstein et al. 1976; Cerella 1979). Being able to acquire natural categories faster is treated as a sign that the animal perceived the pictures of these objects as their real counterparts (confusion mode). However the results can be also explained by that objects in natural categories are often more similar visually to each other than those in artificial categories, therefore simple perceptual learning can produce the observed results.

A second type of tests are called bi-directional discrimination transfer experiments. In these the ability to transfer categorization learnt on pictures to real object and vice versa, can demonstrate that the animal processed the pictures in confusion mode. Additionally if it can also discriminate between the real objects and their pictorial representations, then processing in
equivalence mode can be assumed (e.g. chimpanzees, Pan troglodytes; Savage-Rumbaugh et al. 1980); baboons, Papio anubis; Bovet and Vauclair 1998). The most common critique against this method is that objects and their pictures have a large set of common perceptual features which can aid transfer based purely on perceptual cues.

The next type of picture-object recognition test that relies on acquired responses is referred to as complementary information procedure. It has been employed with e.g. pigeons (Aust and Huber 2006) and long-tailed macaques (Macaca fascicularis; Dasser 1987). The essence of this procedure is that a subject is trained to discriminate a set of incomplete stimuli (e.g. human figures without heads) and then is tested whether the learned discrimination remains when only the previously missing part of the stimuli is available. It is argued that by being able to successfully solve such a task, the subject has to associate the individual parts of real objects from real life experience and transfer this representation to pictures. However being able to achieve this still does not allow us to determine whether the subject processed the pictures in confusion or equivalence mode.

A fourth type of tasks are the cross-modal matching experiments. If the animal can select a picture of an object based on the haptic stimuli of it, than it is assumed that both the haptic stimuli and the picture activated the representation of that particular object (e.g. chimpanzees; Davenport et al. 1975; rhesus monkeys; Tolan et al. 1981). However in these experiments it is again not possible to tell whether the subject was able to realize that the picture it sees is not the real object (equivalence mode).

All the above introduced four test types require extensive training, which makes it difficult to draw inferences about the animal’s spontaneous capacities. Moreover none of the so far mentioned paradigms are able to differentiate between confusion and equivalence mode.

A method that was specifically aimed to test for the presence of equivalence mode (referential understanding of images) was developed by DeLoache (1987), who tested children’s referential abilities using a scale model. In later experiments the methodology was expanded to include the use of pictures (DeLoache 1991) and videos (Troseth and DeLoache 1998) as well. In tests like these the subject is presented with a picture or video of a room on which the position of a hidden reward is shown (Poss and Rochat 2003). If the subject can find the reward in the real room based on this demonstration, and without extensive training, then it can be assumed that it was able to connect the content of the picture with its real world referent, thus it is capable of referential understanding. This testing method has no conventional name.
INTRODUCTION

in the scientific literature therefore it will be referred to throughout the thesis as spontaneous information transfer test. Until now, only humans (children from 2-3 years of age; DeLoache and Burns 1994; Troseth and DeLoache 1998) and chimpanzees (Menzel et al. 1978), could successfully solve such tasks.

2. Dogs (*Canis familiaris*)

2.1. Studies of companion dog cognition in the last two decades

Despite the fact that dogs were subjects of a number of studies from the start of systematic animal behavior research, the investigation of their cognitive abilities has only gained momentum at the onset of the 21st century (e.g. Hare and Tomasello 1999; Miklósi et al. 1998; Topál et al. 1997). The rise of interest was mainly driven by the realization that dogs can be an ideal target of comparative cognition research (Miklósi et al. 2004) because dogs are characterized by both a unique evolutionary history and an exceptional place in the human environment.

The domestic dogs we know nowadays originated from wolf populations between 14000 (Olsen 1985) and 100000 years ago (Vilà et al. 1997). Recent analysis of the canine genome suggests that the ancient dog population was separated from wolves 25-50000 years ago (VonHoldt et al. 2010) and this probably happened in East-Asia (Savolainen et al. 2002). Dogs have the longest domestication history and during this time both their appearance and their behavior diverged from their ancestors. What makes dogs specifically interesting for comparative cognition however is that these behavioral changes were driven by a selective pressure for a better integration with human society, leading to a handful of convergent cognitive adaptations that also characterize their human counterparts (Topál et al. 2009b). These cognitive traits include increased ability of cooperation, utilization of social cues, and decreased defensive aggression compared to hand raised wolves (Gácsi et al. 2005). Additionally the (on an evolutionary scale) recent divergence of dogs from their wild relatives, and the presence of numerous closely related wild Canids (e.g. wolves, jackals, coyotes, dingoes) makes dogs an even better subject of comparative studies. These species/subspecies can be used as controls when attempting to disentangle the effects of phylogenetic changes due to ecological constraints or domestication, and effects of behavioral changes during ontogeny.
Most experiments of cognition rely on procedures in which animals have to tolerate handling by humans. This means that only well socialized or enculturated animals can successfully participate in the tests (unless the researchers are able to devise and execute very sophisticated field studies). However in the case of dogs such special handling is not necessary as dogs’ natural environment is the human social group, since during domestication they have been selected to mature and function in the framework of this hetero-specific environment (Miklósi and Topál 2012). Moreover there is evidence that dogs form functionally similar social bonds with their owners as human infants do with their caregivers (Topál et al. 1998), which is not the case with hand reared wolves (Topál et al. 2005a). Dogs also have a propensity to establish eye contact with humans and, unlike the wolves (Miklósi et al. 2003), they utilize human gaze cues in a number of situations: assessing human’s attention in object retrieval tasks (Bräuer et al. 2004; Call et al. 2003), begging for food (Gácsi et al. 2004), fetching (Kaminski et al. 2009a), commanding situation (Virányi et al. 2004). Dogs also pay more attention to humans with whom they have a close relationship than to other familiar persons (Horn et al. 2013).

This attentiveness towards humans makes dogs, among other cognitive tasks, excellent subjects of social learning experiments. For example, dogs can understand a variety of pointing gestures in a two choice task (Hare and Tomasello 1999; McKinley and Sambrook 2000; Miklósi et al. 1998) and their performance significantly improves in a detour task after observing a human (Pongrácz et al. 2001) or conspecific demonstrator (Pongrácz et al. 2004). Dogs tend to follow social rules (Kubinyi et al. 2003b; Watson et al. 2001) and are capable of social anticipation (Kubinyi et al. 2003a). They show a wide range of social learning skills (for a review see Kubinyi et al. 2009) and they are also sensitive to stimulus and local enhancement (Prato-Previde et al. 2008). There have also been studies that suggest that dogs are to some extent capable of imitation (Miller et al. 2009b; Range et al. 2007; Topál et al. 2006) and that they tend to prefer an inefficient behavior if it was demonstrated by a human over an efficient one (Kupán et al. 2011; Szetei et al. 2003). Dogs also show a marked sensitivity to ostensive cues that signal communicative intention. Similarly to infants but contrary to wolves (Topál et al. 2008) ostensive-communicative cues guide dogs’ attention (Téglás et al. 2012) and influence their performance in object search tasks (Topál et al. 2009a).

The heightened sensitivity to human cueing that makes dogs excellent subjects of social learning experiments also presents a challenge. For example, their attentiveness to various human signals can result in dogs being unintentionally influenced in experiments thus leading
INTRODUCTION

to biased results (Clever Hans effect). Albeit it has been recently shown that dogs are not easily
influenced by unintentional human cueing in some contexts (Pongrácz et al. 2013; Schmidjell
et al. 2012) the vast literature demonstrating dogs’ sensitivity to human attentional cues advises
cautions.

One method to prevent such potential adverse effects by cueing is to use artificial
stimuli. As it has been described in Chapter 1.2, with using images it is possible to standardize
stimulus presentation across subjects, to control for every detail of the stimulus prior of the
experiment and to limit interaction between the subject and the demonstrator if the latter is
presented on a pre-recorded video. The next chapter introduces some of the experiments that
have already successfully used image presentation techniques in studies of dog cognition.

2.2. Studies of dog cognition utilizing images

There are already a number of studies that have employed artificial visual stimuli in the
investigation of dogs’ cognitive abilities (Table 1). Next I will present a handful of them in
more detail to exemplify the different visual presentation methods employed so far.

The earliest of these studies was conducted by Fox in 1971 who used a life-sized
painting of a dog. He found that dogs react similarly to the painting as to an unknown
conspecifics, by sniffing the same body regions (ear and anal areas).

A more recent study demonstrated that using life-sized projected videos of humans can
be used reliably for signaling in a pointing task (Pongrácz et al. 2003). In this study the authors
compared dogs’ performance in a two-way choice task. In the 3D condition the pointing was
performed by a real human, and in the 2D condition a life-sized projected video of a human
gave the pointing cues. In both groups dogs chose the indicated container significantly above
chance level, and their performance did not differ significantly between the two conditions. In
this experiment the projected video was presented in an interactive way: the human was
standing in an adjacent room in front of a camera, and he/she was able to see the reactions of
the dog via a TV screen. This allowed the projected human to respond to the reactions and
attention state of the dogs. However this interactive stimulus presentation did not ensure the
uniformity of the presented stimulus across trials.
<table>
<thead>
<tr>
<th>Study</th>
<th>Stimulus</th>
<th>Phenomenon</th>
<th>Response variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fox (1971)</td>
<td>life-sized panting</td>
<td>intraspecific social behavior</td>
<td>sniffing of picture regions</td>
</tr>
<tr>
<td>Pongrácz et al. (2003)</td>
<td>life-sized projection of live video feed</td>
<td>pointing task</td>
<td>correct choices</td>
</tr>
<tr>
<td>Adachi et al. (2007)</td>
<td>photos presented on computer screen</td>
<td>cross modal matching</td>
<td>looking time</td>
</tr>
<tr>
<td>Range et al. (2008)</td>
<td>photos on a touch sensitive computer screen</td>
<td>categorization of images</td>
<td>conditioned touching response</td>
</tr>
<tr>
<td>Harr et al. (2009)</td>
<td>pre-recorded video on a computer screen</td>
<td>contagious yawning</td>
<td>number of yawns occurring during the presentation</td>
</tr>
<tr>
<td>Kaminski et al. (2009b)</td>
<td>photographs</td>
<td>comprehension of iconic signs</td>
<td>fetching the matching object</td>
</tr>
<tr>
<td>Faragó et al. (2010)</td>
<td>life-sized projected photos</td>
<td>cross modal matching</td>
<td>first look &amp; looking time</td>
</tr>
<tr>
<td>Racca et al. (2010)</td>
<td>projected photos</td>
<td>discrimination of novel and familiar human or dog faces</td>
<td>looking time</td>
</tr>
<tr>
<td>Siniscalchi et al. (2010)</td>
<td>two silhouette drawings</td>
<td>lateralization</td>
<td>first look latency</td>
</tr>
<tr>
<td>Nagasawa et al. (2011)</td>
<td>photographs</td>
<td>discrimination of human facial expressions</td>
<td>conditioned touching response</td>
</tr>
<tr>
<td>Somppi et al. (2012)</td>
<td>photos presented on computer screen</td>
<td>proof of eye tracking method</td>
<td>looking time</td>
</tr>
<tr>
<td>Téglás et al. (2012)</td>
<td>pre-recorded videos on a computer screen</td>
<td>gaze following in communicative and non-communicative contexts</td>
<td>first look &amp; looking time</td>
</tr>
<tr>
<td>Autier-Dérian et al. (2013)</td>
<td>photos presented on computer screen</td>
<td>discrimination of dogs from other species</td>
<td>conditioned touching response</td>
</tr>
</tbody>
</table>

Table 1: List of experiments studying dog cognition that have used images as artificial visual stimuli, presenting the key details of the experiments.

A similar life-sized projection method was used by Faragó and colleagues (2010) but with photos instead of videos. Dogs first heard a playback of an unfamiliar dog’s agonistic (food-guarding) growl and then were presented with two dog pictures (one of the pictures
INTRODUCTION

showed a dog equal in size to the growling dog and one was 30% larger or smaller). The researchers found that dogs looked more on the picture with the correct size in this situation and argued that dogs were able to extract size information from the growls and could match this with the visually acquired information.

The common attribute of the previously introduced three studies was that all of them used life-sized images, with the explicit aim to substitute natural stimuli. Using artificial stimuli in these experiments was motivated by standardizing the presented stimuli, limiting interaction between the displayed human/dog and the subject, and to limit presentation to the visual domain (e.g. by excluding odor cues). The studies described next are similar in respect that they specifically aim to use images as replacement for natural stimuli, however they either use smaller than life sizes or display only partial stimuli (e.g. only the face of a human).

In their studies Racca (2010) and Nagasawa (2011) aimed to investigate whether dogs are able to distinguish different human faces and human facial expressions by using pictures of faces as stimuli. Racca and colleagues (2010) found that dogs look longer on unfamiliar human faces compared to familiar ones, while they tend to look more on familiar dog faces compared to unfamiliar ones. Dogs also looked more on unfamiliar objects than familiar ones, suggesting that they process dog faces differently from human faces as any other visual stimulus in their environment. In a study with similar method Autier-Dérian and colleagues (2013) found that dogs are able to distinguish between faces (heads) of dogs and other animal species.

Using the violation of expectation paradigm Adachi and colleagues (2007) investigated whether dogs can connect the voice of their owners to a picture of their faces. Dogs heard either the voice of a stranger or their owners and were presented with pictures of the one or the other. Dogs tended to look longer on the pictures when they heard a mismatching voice beforehand, demonstrating that this constellation did not correspond to their expectations.

Using pre-recorded videos clips Harr and colleagues (2009) tested whether dogs would respond to yawns of humans, as it was found by an earlier study that used live demonstration of yawning (Joly-Mascheroni et al. 2008). Interestingly they were unable to replicate the results of the previous study as dogs did not show increased rates of yawning when they were watching the videos.

A method borrowed from human infant studies has also been successfully employed recently with dogs. By using dedicated eye-tracking devices it became possible to precisely
track the target of dogs’ attention. The method has been demonstrated to work successfully both by using pictures (Somppi et al. 2012) and video clips (Téglás et al. 2012).

In contrast to the studies presented so far the following two studies used abstract artificial visual stimuli or realistic stimuli but in an abstract task. Siniscalchi and colleagues (2010) used black and white silhouettes of animals (a dog, a cat and a snake figure) to test dogs’ looking responses to the sudden appearance of these stimuli. Results indicated that dogs react quicker to cat and snake pictures if they are presented on the left side, however no such difference was found in regard to the dog pictures.

Range and colleagues (2008) used multiple photographs of dogs and landscapes presented on a computer screen in their touch-screen categorization experiment. All dogs were able to learn this categorization needing on average 900 trials. However it is very likely that dogs treated the pictures as abstract visual stimuli and did not learn to categorize them based on their content because: a) an extensive training preceded the tests that involved abstract geometric figures; b) dogs’ eyes are not able to focus well on stimuli as close as the screen was in this experiment (see in next chapter).

These lastly raised issues lead to the questions of how dogs perceive visual stimuli in general and how they process artificial stimuli in particular. Chapter 2.3. will show what we know about dogs’ visual perception based on the literature to date, and Chapter 2.4. will discuss what can be inferred about dogs’ visual processing based on the studies presented in the current chapter.

2.3. The visual perception of dogs

Compared to their excellent sense of smell (Neuhaus 1953; Quignon et al. 2003) and their sense of hearing (Heffner 1983; Lipman and Grassi 1942) which both surpass that of humans’, dogs’ sense of vision is relatively poorer (Miller and Murphy 1995). Still many of the experiments conducted with dogs (not limited to those utilizing artificial visual stimuli) are designed to test abilities that rely primarily on vision and require recognizing or making distinctions between fine details of the visual domain. Such experiments include visual displacement tasks (Collier-Baker et al. 2004; Fiset et al. 2006; Gagnon and Doré 1992), pointing tasks (McKinley and Sambrook 2000; Soproni et al. 2001), tasks about gaze following (Hare and Tomasello 1999; Kaminski et al. 2009a), facial discrimination tasks (Adachi et al.
INTRODUCTION

2007; Nagasawa et al. 2011), studies of attention (Siniscalchi et al. 2010), studies with touch-screen (Range et al. 2008) or any other tasks involving visual stimulus presentation (Kaminski et al. 2009b; Pongrácz et al. 2003).

In none of the above listed publications do the authors explicitly state that they have designed the experiments taking in account the differences between the visual perception of humans and dogs. However, any unaccounted difference between the visual perception of dogs and humans could mean that dogs perceive the visual stimuli in a way not anticipated by the researchers; which consequently could lead to puzzling results or to a misinterpretation of dogs’ behavior in the tests. Therefore it is important to assess how dogs would perceive visual stimuli and this is especially true when using artificial visual stimuli. The following section will present an overview of the differences between the visual perception of dogs and humans by starting with the discussion of color perception.

It is known that dogs have a dichromatic color vision (Jacobs et al. 1993a; Neitz et al. 1989), meaning that dogs have two types of light sensitive photopigments in the cone cells of their retina. Humans on the other hand have three different types of photopigments and therefore they can distinguish more hues than dogs can. The three photopigments of humans are often referred to as red, green and blue according to the apparent hue of light they are most sensitive to. The two photopigments of dogs, based on their sensitivity range, are similar to the human green and blue pigments. According to this, dogs’ color vision must be similar to that of a person suffering from deuteranopia (a type of red-green colorblindness). For instance dogs have difficulties discriminating red from yellow or green, and the colors that humans see as greenish-blue may be perceived by dogs as being achromatic (white or gray).

The next difference discussed in this chapter is brightness discrimination. Brightness discrimination is the ability to differentiate between various shades. It is measured by determining the smallest discernible difference in brightness between two stimuli, and is expressed as the ratio of the intensity (Weber fraction). The Weber fraction for humans is 0.11 (Griebel and Schmid 1997) whereas the Weber fraction for dogs is 0.22 (Pretterer et al. 2004). Based on these studies dogs’ brightness discrimination is two times poorer than that of humans.

The third difference in visual perception between dogs and humans is acuity. Visual acuity is a measure of the spatial resolution of the visual system. It is often measured in cycles per degree (CPD), which is the number of cycles of a grating (dark and light bands) subtended at the eye per degree. The maximum visual acuity is equal to the highest CPD value where the
INTRODUCTION

grating is still perceived as being comprised of separate bands. The maximum visual acuity of the human eye is around 50 CPD (Russ 2006) and 60 CPD (Campbell and Green 1965). The measurements of dogs' visual acuity vary between 7.5-9 CPD (Ofri et al. 1997) and 11.6 CPD (Odom et al. 1983). According to these measurements dogs' visual acuity is four to eight times worse than that of humans. Another factor affecting visual acuity is the eye’s ability to focus (accommodate) on objects at various distances. The accommodative ranges of dogs’ is 2 to 3 diopters which means that the nearest distance they are able to focus to is 33 or 50cm (Miller and Murphy 1995). As a comparison humans can focus on objects as close as 10cm (Blaker 1980).

By taking in account the above listed differences in the visual perception of dogs and humans, researchers can design experiments by selecting more appropriate visual stimuli for their subjects (e.g. selecting appropriate colors when discrimination of stimuli is required; placing stimuli further than 50cm from the dog; ensuring that important details are large enough for dogs to perceive easily). However, seeing the effects of these differences directly on the actual stimuli or on the experimental environment could be even more helpful in the designing phase of a study. This is possible with the help of an image processing tool, which demonstrates the difference in the visual perception the dogs and humans. Such a web based tool was created by the author and can be found at the web address 'dog-vision.com’.

This image processing tool enables users to observe the effects of dichromacy, decreased brightness discrimination and decreased visual acuity, separately or arbitrarily combined. It requires only a web browser for functioning and it is available free of charge. The tool enables users to upload digital images, which are processed in to display how dogs would perceive them. The modified images can be viewed on the webpage or users can download them for later use. An example image processed by the tool can be seen on Figure 1, details about the functioning of the processing algorithm can be found on the webpage. Naturally such a tool cannot show precisely how dogs perceive these images, but it can demonstrate the differences compared to how humans’ would perceive the same scene.
Next to the static factors of vision introduced so far there are also dynamic factors. These are relevant if moving stimuli are used (e.g. videos or animations). Dogs’ sensitivity to motion is regarded to be better than that of humans (Miller and Murphy 1995), however there are no explicit measurements comparing the two. There is more information available however about the flicker sensitivity of dogs. Dogs’ maximal critical flicker fusion frequency lies at 70Hz (Coile et al. 1989) which is slightly higher than the 60Hz of humans’ (Landis 1954).

The practical consequence of dogs higher CFF is that image presenting devices that were designed with a fast enough flicker ratio for human observers, might show a perceptible flicker for dogs. Typically older cathode ray tube television screens and computer monitors fall into this category which refreshed the whole image between every frame. Modern LCD monitors and projectors function on the basis of different principles and they only change the
INTRODUCTION

differing portions of an image between frames. Additionally flicker sensitivity is lower in low light conditions (Porter 1902) which are typical in most stimulus presentation scenarios. This effect can be grasped by considering that the maximum human CFF is 60Hz however cinema films are presented with 24Hz but still appear non-flickering for most humans.

What we can see from comparing dogs’ visual system to that of humans is that the image presentation devices tuned to the human eye would produce adequate images for dogs’ visual systems under most circumstances. However knowing how dogs’ visual system perceives images does not tell us how they process cognitively the presentations. To investigate this question the next chapter will summarize what conclusions can be drawn about dogs’ image processing abilities based on studies using artificial visual stimuli.

2.4. What can be inferred about dogs’ image processing from studies to date

The many successful applications of images in the studies listed earlier and the comparison of dogs’ visual system with the one of humans’ may suggests that dogs have little difficulty perceiving the content of pictures and videos. However the question how dogs process these stimuli has not been raised in any of experiments to date, therefore this chapter will summarize what conclusions can be drawn from the results of already published studies. The analysis will attempt to determine in which of the three processing modes proposed by Fagot (2010) dogs perceived the images; and will be limited to experiments that did not involve conditioned responses, since drawing inferences from such studies without adequate controls can be misleading as it has been shown in Chapter 1.5.

The study of (Fox 1971) where the author observed that dogs behaved similarly to a painting as they would done seeing a real conspecific provides an excellent starting point. The obvious conclusion from dogs’ reaction is that they perceived the life-sized painting in confusion mode since the very definition of confusion mode is: “reacting the same way to an image as to the real object”.

In other studies that used images to substitute for natural stimuli (Adachi et al. 2007; Faragó et al. 2010; Racca et al. 2010; Somppi et al. 2012; Téglás et al. 2012) dogs reacted to the stimuli as was expected by the researchers. This confirms that artificial stimulus presentations with current methods are quite robust, and that dogs perceive images at least in confusion mode and not in independence mode.
However in the study of Harr and colleagues (2009) dogs did not respond to the recorded videos of human yawns as they did in an earlier experiment with real demonstrators (Joly-Mascheroni et al. 2008). The lack of a positive result in this study could be attributed to specific technical problems with the presentation (a commercial laptop screen was used) as under similar conditions (video presentation of a human on a computer screen) dogs were able to respond appropriately (Téglás et al. 2012). However without further studies on contagious yawning or on the limits of smaller than life-sized video presentations it is not possible to determine the exact reason why dogs failed to react to the yawns on the videos.

The study of Siniscalchi and colleagues (2010) differs in a critical point from the previous experiments. Here dogs were presented only with silhouettes of animals, yet they still reacted differently to the stimuli, showing that they could differentiate between them. However this experiment was unlike from all the others insofar that the two stimuli were always presented laterally of the dogs. Therefore they perceived them on the periphery of their vision where visual acuity is lower. So even if realistic pictures would have been presented for dogs in this experiment they likely still would have not perceived them having more details than a silhouette.

The experiment of Pongrácz and colleagues (2003) provides the most details that enable theorizing about the image processing abilities of dogs. In their experiment the researchers found that dogs could follow the pointing gestures of a human equally well when he was projected as a life-sized video as when he was present in the room. Pongrácz and colleagues (2003) also found that dogs’ performance did not improve during the trials in the video condition, therefore they ruled out the possibility of learning during the experiment. Based on this later result it is possible to exclude independence mode since if dogs would have only perceived the video as a set of simple unrelated features but not as a human, then they would have required at least some degree of learning to solve the pointing task. This still leaves the other two options open (confusion and equivalence mode) but it is not possible to exclude the one or the other.

However there was a second experiment in this study where dogs had to follow the commands issued by their owners. Either when their owner was present (3D condition) or when the owner was only visible via a life-sized projected video (2D condition: the owner’s voice was also transmitted via a loudspeaker behind the projector canvas). The results showed that dogs followed the commands less in the 2D condition than in the 3D condition. This result is intriguing as it does not align completely with what one would expect if dogs would have
INTRODUCTION

processed the videos in confusion mode. Performing lower in the 2D condition could be interpreted as dogs realizing that the owner was not present in the room, which could mean that they processed the video in equivalence mode. However differences were found only for certain commands and showing a lower performance in the 2D condition is not clear proof of equivalence mode processing.

Nevertheless dogs live among humans and are constantly exposed to referential artifacts that inhabit the human world (e.g.: pictures, television, mirrors) and anecdotal evidence from owners also suggests that dogs react appropriately to these artifacts. It has been also shown that humans require some degree of exposure to images to be able to correctly perceive and interpret them. For e.g. Miller (1973) demonstrated that humans who have never seen pictures have difficulty recognizing what is represented in black-and-white photographs. Therefore it is not unreasonable to expect that via this heightened exposure dogs could become able to realize the referential nature of images and process them in equivalence mode. There are also a number of studies that demonstrate dogs showing referential capabilities in different domains which will be introduced in the following chapter.

2.5. Studies focused on the referential abilities of dogs

Referential understanding in the broad sense is the ability to perceive an object (e.g.: picture, video, replica, scale model) or an action (e.g.: verbal label, pointing) as being content-full and as standing for other entities in the world (DeLoache 1991; Gliga and Csibra 2009). In the field of animal communication researchers coined the term ‘functionally referential signal’. The term is used for animal signals that seem to refer to external phenomena and not being only a reflection of the internal state of the animal. However by calling these signals ‘functional’ researchers forego the philosophical arguments on the problem of ‘meaning’ and acknowledge the constraints inherent in the study of nonhuman communication systems (Macedonia and Evans 2010). Therefore the referential understanding abilities of animals are also often denoted as being ‘functional’ to emphasize that no detailed assumption is made about the underlying cognitive processes, and that these cognitive processes are not regarded as being equivalent to how humans would process the same signals/stimuli. There are three separate fields where to date dogs have shown to be able to comprehend referential signals: a) understanding pointing gestures, b) using a large vocabulary of verbal labels, c) comprehending iconic signs.
INTRODUCTION

Several studies show that dogs can easily follow many forms of human pointing gestures (Agnetta et al. 2000; Soproni et al. 2001), and can even follow directional cues of conspecifics (Hare and Tomasello 1999), consequently some researchers argue that dogs understand the referential nature of pointing (Lakatos et al. 2009; Scheider et al. 2011). In contrasts others emphasize the role of experience during ontogeny, and the role of associative learning in the emergence of this capacity (Elgier et al. 2012). Interestingly this capacity in dogs seems to be symmetrical as dogs also are able to inform humans about the location of objects. In the experiment of Miklósi and colleagues (2000) dogs displayed showing behavior towards the location of hidden food, but only if the human was present in the room. The authors argued that these findings provide evidence that dogs are capable of functionally referential communication.

A few highly trained dogs can fetch a named object with high reliability, and with using a vocabulary of up to thousand words. According to the authors of the studies this suggests that dogs understand verbal object labels in a referential manner (Kaminski et al. 2004; Pilley and Reid 2011). Counter arguments generally point out that contrary to humans, dogs probably understand these labels in an associative way. Also humans can use object labels in a wide variety of different contexts but for dogs these labels are rather only commands to fetch particular objects (Bloom 2004; Markman and Abelev 2004). However these arguments still do not mean that this capacity of dogs would not fall under the wider umbrella of functional referentiality.

In their experiment Kaminski and colleagues (2009b) demonstrated that dogs are able to use iconic signs (life sized replicas, miniature replicas or photographs) to correctly retrieve the corresponding object from a pool of objects. The authors of this paper argue that mastering this ability without previous training is a proof of that dogs understood the referential nature of the iconic signs. However fewer dogs were able to use photographs than replicas to retrieve the objects, which suggests that making connection between the artificial stimuli and their real life referent was cognitively more demanding for dogs.
**STUDIES**

As it was shown in Chapter 2.2. there are already several dog cognition studies that employ various forms of images as artificial visual stimuli, and the number of such studies is expected to increase in the future. The growing popularity of this technique can be attributed to the three main benefits of this method: repeatability, stimulus control and ease of use (see Chapter 1.2.). The analysis of dogs’ visual system (Chapter 2.3.) and the success of these studies suggests that images are adequate replacements of natural stimuli in cognitive experiments with dogs. The experiments so far indicate that dogs process images either in confusion or equivalence mode, however none of the studies to date were specifically designed to answer this question. This gap in our knowledge can become a serious problem as the interpretation of any experiment using images largely depends on how dogs processed these stimuli.

Therefore the central aim of this thesis is to investigate how dogs process images and to explore some of the methods that could be used to answer this question. There is a wide variety of methods available to test image processing in animals (Chapter 1.4). In the frame of this thesis three different methods were employed: a) a spontaneous information transfer test (Study 2), b) a spontaneous response test based on the violation of expectation paradigm (Study 3), c) and an acquired response test using the touch-screen testing procedure (Study 4). The detailed description of each study’s method and theoretical background can be found in the corresponding chapters and only a short overview of the most important methodological parameters is presented here.

The spontaneous information transfer test was based on the method of Troseth and DeLoache (1998). In our test dogs observed a pre-recorded, life-sized projected video of a human hiding an object, and then were allowed to search for this object in the room depicted on the video. In the violation of expectation test dogs were also shown videos of a human hiding an object, however this time their reactions to ‘possible’ and ‘impossible’ scenes was recorded.

2 The violation of expectation paradigm was pioneered by Baillargeon and colleagues (1985) and it was initially used to test the object permanence ability (the ability to understanding that objects continue to exist even when they are no longer available to immediate perception in the environment; Piaget, 1954) of 5 month old infants. The gist of this experimental method is that subjects are first habituated to a certain event. Next they are presented with a ‘possible’ or an ‘impossible’ variant of the original scene (one contradicting the rule of object permanence) and their reaction (typically looking time) is recorded. “On the basis of the commonly-held assumption that infants react to novel or surprising events with prolonged attention” (Baillargeon et al. 1985) the authors expected that infants would look longer at the ‘impossible’ scene if what they saw would violate their expectation formed via their object permanence abilities. On the other hand if they did not yet possess such ability, they would not display different amount looking towards the two test scenes.
STUDIES

The touch-screen test in this thesis differs from the ones described in Chapter 1.5, where subjects learned to categorize or discriminate a set of images. In our test dogs were shown computer generated animations that resembled a visual displacement task, and they had to choose the location where an animated target stimulus disappeared.

All experiments relied on the visual displacement task as their underlying test. The inclusion of a visual displacement task is evident in case of the spontaneous information transfer test as this method requires subjects to find a hidden object. We choose to include this method into the violation of expectation test to have a task which motivates dogs to pay attention to the demonstration for multiple trials. In the touch-screen test our aim was to compare dogs’ performance in this abstract 2D task with their real world performance because there is an ample amount of information available about how dogs perform in visual displacement tasks.

Both the spontaneous information transfer test and the violation of expectation test used pre-recorded, life-sized videos (with sound) of a human demonstrator as stimuli. This choice was made because life sized videos were already used successfully in Pongrác and colleagues’ (2003) study and because their results suggested that dogs process the videos in equivalence mode. In the touch-screen test we employed computer generated animations as stimuli because life-sized videos would have not been possible to present due to the size limitations of the touch-screen.

A key feature of the videos in spontaneous information transfer test and the violation of expectation test is that a human is the demonstrator. To ensure that dogs’ attention is directed to the video the human also addresses the dog verbally. However it is known that a communicative (ostensive) context affects dogs’ performance in visual displacement tasks, especially if multiple trials or repetition are involved. Topál and colleagues (2009a) found that in an ostensive-communicative context dogs tend to choose the location where an object was found in the previous trials instead of the location where they have seen it hidden in the current

3 Visual displacement tasks were primarily employed in experiments that studied dogs’ object permanence abilities and these studies form one of the earliest branch of methodological dog cognition research (e.g. Gagnon and Doré 1992; Triana and Pasnak 1981). In the simplest form of this task the subject observes as a target object (e.g. a toy or a treat) is being moved to one of several hiding locations. After the target object disappears at a certain location the subject is allowed to search for the object. Being able to reliably locate the object in a simple or more complex version of the task (e.g. with a delay before search) is regarded as proof that the subject possesses a certain level of object permanence ability (see: Piaget 1954). Dogs have demonstrated their ability to solve even the most elaborate visible displacement tasks (Gagnon and Doré 1993).
trial (this type of choice pattern is referred to as ‘perseverative error’). This tendency of dogs could also have a profound effect on their performance in our study, because we plan to run multiple trials with hiding an object multiple times to the same location. Since the methodology of Topál and colleagues (2009a) had some critical details which made it difficult to determine how their findings would influence our studies, we planned to repeat their experiment with a simplified method (Study 1).
STUDY 1: The effect of ostensive cues on perseverative errors in a visible displacement task

Aims and questions

The term ‘perseverative error’ is used when a subject engaged in a string of repetitive tasks involving more than one option, continues to perform the previously successful solution even though it is erroneous in the actual context. This phenomenon is known as ‘A-not-B error’ when referring to a two choice object search task. Here the perseverative error manifests when the subject searches at the previously visited and rewarded A location, while the object has clearly been placed to a second B location.

In their study, Topál and colleagues (2009a), found that dogs, but not wolves, were prone to commit A-not-B errors in an ostensive-communicative context when a human experimenter repeatedly hid a target object. The fact that intensively socialized wolves did not commit perseverative errors in the presence of ostensive cues was regarded as proof that dogs have a special sensitivity to these kinds of communicative signals, which are otherwise often adaptively used in human-human (Csibra and Gergely 2009) and human-dog interactions (e.g. Pongrácz et al. 2004).

The research of A-not B errors originates from studies conducted with young children. Infants up to 12 months of age were found to commit A-not-B errors (see Marcovitch and Zelazo (1999) for a review). Originally these errors were attributed to the not yet developed object permanence ability of this age group (Piaget 1954). During the past decades, additional (and partially overlapping) explanations have been proposed regarding the source of the phenomenon: (i) deficit in the inhibitory control of reaching towards the A location (Diamond 1985), (ii) constraints on short-term memory (Bjork and Cummings 1984), (iii) attention bias towards the A location (Ruffman and Langman 2002), (iv) covert imitation of the reaching towards location A through the activation of the mirror neuron system (Longo and Bertenthal 2006).

In an earlier experiment Topál and colleagues (2008), found that 10 month old infants commit perseverative errors only if the hiding event was performed in an ostensive-communicative context. The authors concluded that ostensive-communicative signals induce a receptive learning attitude in infants who are generally sensitive to the ‘teaching aspect’ of the actions of adults (Csibra and Gergely 2006; Gergely et al. 2007). This heightened willingness of infants to learn in a social context is referred to as ‘natural pedagogy’ (Csibra and Gergely
As a consequence, infants tend to (mis)interpret the repeated (and ostensively reinforced) hiding events at location A as manifestations of generalizable knowledge (‘This kind of object is to be found in container A’) and/or demonstrations of a particular desired behavior (‘Search here!’).

These findings (Topál et al. 2009a; Topál et al. 2008) have triggered a vivid theoretical debate about possible alternative explanations why children, dogs and wolves may or may not show perseverative tendencies in particular social and non-social scenarios of the A-not-B hiding protocol. For example, Fiset (2010) in his commentary suggested that dogs show perseveration because they have limited attention span, while wolves are better in focusing on moving targets. However, there is no actual proof of higher attention span in wolves than dogs, and wolves are reportedly less skillful in following human given cues (e.g. Gácsi et al. 2009; Miklósi et al. 2003).

Given that the target object is moved behind location A prior to being placed to the B location in the B trials of Topál and colleagues (2009a) experiment, Fiset (2010) also suggested that these sham baitings can lead to perseveration because the subject may think erroneously that the object was left at location A. Topál and colleagues (2010) showed that this was not the case, because dogs kept on persevering even if the experimenter did not bend down (‘sham-bait’) at location A while she was moving towards location B.

In addition, Marshall-Pescini and colleagues (2010) pointed out that the error may stem from the ‘unbalanced’ presentation of attention getting signals. In fact, according to the protocol (Topál et al. 2009a; Topál et al. 2008), the experimenter used ostensive cues only during the hiding events to location A and subjects’ attentional bias towards these local enhancement cues could have led to perseveration. Topál and colleagues (2010) found however, that dogs kept on persevering even after hiding the toy at location B was reinforced with non-verbal acoustic cues (squeezing a rubber toy). This raises the possibility that it was not the unbalanced acoustic reinforcement itself, but rather the specific effect of the expression of the experimenter’s communicative intent, which may be a key factor of dogs’ tendency to commit perseverative errors. We should note however, that ostensive addressing (calling a dogs’ name) may be a more powerful cue for grabbing attention than the sound of a toy. Although in a more recent study, using a slightly modified protocol of Topál and colleagues (2009a), Kis and colleagues (2012) provided further support for the significant effect of socio-communicative cues in the emergence of A-not-B error in dogs, these studies did not directly test whether ostensive and non-ostensive cues elicited a comparable level of attention.
Summarizing the critics of several researchers and the counter-arguments of Topál and colleagues (2010), it is clear that the original method (Topál et al. 2009a; Topál et al. 2008) possessed particular features that did not enable to determine unequivocally the factors underlying the emergence of perseverative errors in dogs and infants. Therefore the question still remained open whether dogs or infants are influenced by ostensive communication in a simple visible displacement task: a) where ostensive cues are given equally in all trials; b) that lacks ‘sham-baiting’; c) and has a minimal number of A trials (only 2 opposed to 4 or more employed in previous studies). Next to its contribution to our understanding of the cognitive mechanisms that underlie A-not-B errors, finding the answer to this question is also highly relevant for studies utilizing visual displacement with multiple trials in an ostensive context, including the next two studies (2 and 3) of this thesis as well.

Therefore with the intention of making a ‘more classic’ approach (c.f. Piaget, 1954) to investigate the role that ostensive communication plays in eliciting perseverative search error, we devised a two-way visible displacement task. Unlike the method used in the previous studies (Topál et al. 2010; Topál et al. 2009a; Topál et al. 2008), our object hiding procedure was characterized by high levels of symmetry: the experimenter used the same kind and amount of ostensive cueing during each trial and the manipulations of location A (sham-baitings) during B-trials were omitted. During the test, the object was first hidden in plain view of the subject at location A (trial A1), this was repeated once more (trial A2), after which the object was hidden at location B in the same direct manner (trial B1). The B-trial was also repeated (trial B2), and finally the object was hidden to the first hiding spot again (trial A3). Subjects were given the opportunity to search for the object in each trial.

Next to a group of adult companion dogs we also tested a group of two year old children. Our question was whether dogs and children would still commit perseverative errors in an ostensive-communicative hiding context, if the sole (asymmetric) factor to elicit this was the repeated hiding to location A in the first two trials.

Adult dogs and two year old children have fully developed understanding of object permanence (Anderson 1955; Gagnon and Doré 1994) and the task constitutes a simple, single visible displacement problem. One can hypothesize therefore that under these circumstances dogs and children will not show perseverative search bias. Such a finding would mean that the perseverative errors in earlier studies were likely induced by the combination of unbalanced presentation of ostensive cues and the sham baiting of location A during B trials. However if we find a significant bias toward selecting location A in the first B trial, that would mean that
perseverative errors are linked to human ostensive communication, since children of this age group and dogs do not show perseverative errors in a non-ostensive context. (Sophian and Wellman 1983; Topál et al. 2009a). More specifically, this finding would confirm that perseverative errors are caused by social facilitation in the dog and a tendency in children to misinterpret the hiding action at location A as a potential teaching demonstration as was suggested by earlier studies (Kis et al. 2012; Topál et al. 2010; Topál et al. 2009c; Topál et al. 2009a; Topál et al. 2008).

Methods

This experiment consisted of two experimental groups. The principal arrangement and test procedure was the same in both groups. Two hiding locations were used (A and B) and the test consisted of five trials. In trials 1 and 2 the target object was placed to location A. In trials 3 and 4 the object was placed to location B. In trial 5 the object was again placed to location A. For half of the subjects hiding location A was on the left, for the other half it was on the right. All tests were recorded with camcorders, the video footage was digitalized and analyzed later.

Group 1: Dogs (Canis familiaris)

Subjects

Dogs ($N = 20$; 9 females and 11 males) and their owners were recruited from clients of various dog training schools and participants of dog-competitions. Owners (O) were instructed in advance how to behave and what to do during the test, however they were unaware of the hypothesis of the study. Only dogs older than 1 year were tested ($M = 2.8, SD = 1.5$), and various breeds were included (1 Airedale Terrier, 4 Border Collies, 2 German Shepherd Dogs, 2 Golden Retrievers, 1 Groenendael, 1 Vizsla, 1 Labrador Retriever, 2 Miniature Poodles, 1 mixed breed, 1 Mudi, 4 Belgian Tervurens). The dogs had to be motivated to retrieve a ball or rubber toy. This was tested a priori to the experiment by throwing a toy three times and observing whether the dog retrieved it.

Apparatus

The hiding locations (2 plastic panels, 40cm×40cm with two bent 10cm extensions on each side) were placed 2m from each other and 15m from the starting position (SP). A plastic flower pot (diameter 16cm) was placed behind each panel. The pots served for holding the target object during the trials, and ensured that the dogs had to look into one of them to find the
object. We used a rubber squeezable toy or a tennis ball as the target, depending on the preference of the dog. All tests were performed outdoors at the same Hungarian dog training school.

**Procedure**

The procedure was the same in both the A and B trials. Each trial started with the O, the dog and the experimenter (E) standing at the SP. The E took the target object, showed it to the dog and walked straight towards the actual hiding location without passing behind the other screen. While walking, the E maintained the dog’s attention on himself by frequently turning his head toward the dog, calling the dog’s name and waving the target object. After reaching the actual location, the E placed the object behind the plastic panel into the pot, then displayed his empty hands and returned to the SP. While walking back the E continued talking to the dog and displaying his empty hands. After the E returned to the SP, the dog was released and encouraged to search for the target object, however neither the O, nor the E gave any hints about the location of the target.

The average (± SD) time interval between the disappearance of the object and the start of the search was 9.7 ± 1.9s. The dog was allowed to search for the object until finding it. After finding the object the dog was praised by the O and the E, and recalled to the SP. Finally the O took the object from the dog, handed it to the E and the next trial began.

**Group 2: Two year old Children**

**Subjects**

Two year old (± 14 days) children (N =20; 10 boys and 10 girls) and their parents were recruited from a database at the Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences (ICNS HAS). The parents of all participants gave informed consent, however they were unaware of the exact hypothesis of the study.

**Apparatus**

The equipment and its arrangement during the test trials was similar to the ones used by Topál and colleagues (2008). Two identical brown plastic pots served as containers and were placed upside down on a rectangular table 40cm apart. We used one of 4 small toys (a toy train, a plastic cube containing a crocodile and rattling balls, a toy plane and a plush mouse) as target objects based on the child’s preference. Experiments were conducted in a test room (5m x 4m) of the ICNS HAS.
**Procedure**

The experimental procedure included a warm-up and a test phase. During the experiment only the child, the parent and the experimenter (E – in this study Anna Gergely) were present in the room.

**Warm up**

Before the test trials, the parent put the child on a blanket on the floor, and the E attracted the child’s attention to the objects placed there (the 2 plastic pots and the 4 small toys). The child was allowed to explore the objects freely for one minute. Then the E initiated playing with the objects (approx. 1-2 min). This procedure was designed to give a chance to the child to become familiar with the environment, the objects and the E, and to select one of the toys, which could be used as the target in the test trials.

**Test trials**

The child sat on the parent's lap on one side of the table, equidistant (70cm) from the two containers. Before each trial, the parent was instructed to close the eyes during the entire object hiding sequence (Figure 2).

The E, sitting opposite to the child, placed the target object to the SP, between container A and B (20cm from each). The E attracted the child’s attention verbally (“[Name]! + Look!”) and by making a conspicuous noise with the target object (tapping the toy on the table). Then the E started to move the object slowly towards the actual container while talking to the child (“Look at the nice toy, now it moves here.”), lifted the container and then lowered it onto the toy. At this point the E established eye contact with the child and addressed him/her (“Look!”) while looking back and forth twice between the container and the child. During this procedure the E ensured that the child was following the movement of the object and repeated any step that the child was not paying attention to.

After lowering the container, the E waited for 4s before pushing the containers simultaneously towards the child by means of an 80cm long stick. The stick was held horizontally along the table, and it was used for moving the two containers at the same time without the E directly touching them. When the containers arrived within the child’s reach, the E withdrew the stick and lowered it below the table. At this moment the parent was allowed to open the eyes but was not allowed to interact with the child.
The average (± SD) time interval between the hiding (disappearance of the toy) and the start of the search was 8.72 ± 1.13s. The child was allowed to search for the object until finding it. Finding the toy was defined as the child lifting the pot that covered the toy. In some cases the children did not lift the pots, but only touched them and then withdrew their hands. If after touching the pot the child still did not lift the pot after several seconds, the E lifted the pot that was touched by the child and displayed its contents to him/her. After finding the toy the child was allowed to play with it for a short time. During this time, the E moved the containers to their original position. Finally the E took the object from the child, placed the toy to the SP and the next trial began.
Data Collection and Analysis

All trials were video recorded and the recordings coded with Solomon Coder beta (© 2012 by András Péter). For the analyses, we coded subjects’ first choices between the two locations. In group 1 (dogs) the first inspected location was regarded as the subject’s choice and a choice was scored as correct if the dog lowered its head behind the baited plastic panel and looked into the pot behind it. In group 2 (two year old children) three different aspects of the participants’ choice behavior were recorded: looking at, touching and lifting one of the containers. A trial was regarded as correct if the subject (1) looked at, and/or (2) touched and/or (3) lifted the baited pot. These behaviors represent different levels of executive functioning and did not coincide in all trials, therefore we analyzed these behaviors separately. In all groups a trial was terminated when the subject selected and inspected the baited location.

Subjects’ first choices were analyzed with IBM SPSS Statistics 21 and R 3.0.1. All data were tested for normality with Kolmogorov-Smirnov tests with Lilliefors significance correction. According to the results of the tests the data were not normally distributed, therefore we used nonparametric statistical methods. When comparing the number of correct choices to that expected by chance we used binomial tests. To account for the increased chance of type I errors due to multiple comparisons, we adjusted the p values in each test battery using the method by Hochberg (1988) as was described by Wright (1992). The adjusted p values are marked as $p_{Hoch}$. When comparing the number of correct choices between trials in group 1 we used the related-samples Cohran’s Q test. When comparing the number of correct choices between trials and between choice behavior types (look, touch, lift) in group 2 we used a generalized linear mixed model. The model used a binary logistic link, correct/incorrect choices as the target variable, number of trial, type of choice behavior and their interaction as fixed effects. We allowed the degree of freedom to vary between tests due to the relatively moderate sample size. Also to compensate for potential deviations from the model’s assumptions we used robust covariance estimates. The LSD method was used to compensate for multiple comparisons (adjusted p-values are marked as $p_{adj}$).
Results

Group 1: Dogs

Using binomial tests we compared the number of correct and incorrect choices to the ratio expected by chance in each trial (Figure 3). The results showed that dogs first chose the correct location above chance in all the A trials (A1: $p_{Hoch} < .001$; A2 $p_{Hoch} < .001$; A3 $p_{Hoch} < .01$). In contrast, they chose the correct location first significantly below chance in trial B1 ($p_{Hoch} < .05$) whereas in trial B2 the ratio of correct choices did not differ from chance performance ($p_{Hoch} = .503$). When comparing the number of correct choices between trials with a related-samples Cochran’s Q test we found a significant overall difference ($N = 20$, $Q(4) = 42.61$, $p < .001$). The post-hoc pairwise comparison showed that in the first B trial dogs made significantly fewer correct choices than in either of the A trials (Figure 3).

![Figure 3](image-url)

Figure 3 Ratio of correctly choosing individuals (looking into the pot) in group 1. Binomial tests comparing the number of correct choices to the ratio predicted by chance (n.s.: $p_{Hoch} \geq .05$, *: $p_{Hoch} < .05$, **: $p_{Hoch} < .01$, ***: $p_{Hoch} < .001$; dashed horizontal line marks the ratio of correct choices expected by chance (50%)). Related-samples Cochran’s Q test comparing the number of correct choices between trials (capital letters P and Q mark the trials that differ from each other significantly according to the post-hoc pairwise comparison, trials with the same letter do not differ significantly).
**Group 2: Two year old Children**

In the case of two year old children, we analyzed three behaviors indicating choice: looking at, touching and raising one of the two pots. With binomial tests we compared the number of correct and incorrect choices to the ratio expected by chance in each trial for the three choice behaviors (Figure 4). On the basis of their first look children selected the baited location in all A trials as well as in the second B trial significantly more often than expected by chance (A1: $p_{Hoch} < .01$; A2 $p_{Hoch} < .001$; B1 $p_{Hoch} = .26$; B2 $p_{Hoch} < .01$; A3 $p_{Hoch} < .05$). In contrast, the analysis of their more explicit responses (first touch: A1: $p_{Hoch} < .01$; A2 $p_{Hoch} < .001$; B1 $p_{Hoch} = .50$; B2 $p_{Hoch} = .08$; A3 $p_{Hoch} < .01$; visual inspection of the pot: A1: $p_{Hoch} < .001$; A2 $p_{Hoch} < .001$; B1 $p_{Hoch} = .50$; B2 $p_{Hoch} = .08$; A3 $p_{Hoch} < .01$), suggest a somewhat different picture of their performance: subjects did not show a preferential bias towards the baited location in either of the B trials (Figure 4).

We also tested whether the number of correct choices differed between trials, and whether the different choice behaviors yielded different numbers of correct choices. The generalized linear mixed model indicated a significant effect of choice behavior type ($F(2, 33) = 4.46, p < .05$) and number of trial ($F(4, 21) = 6.15, p < .01$) but no significant interaction between choice behavior type and number of trial ($F(7, 285) = 1.87, p = .074$). Calculating the pairwise contrasts for the estimated means of the choice behavior types yielded no significant differences (look vs. touch: $t(20) = 0.54, p_{adj} = .122$; look vs. lift: $t(28) = 1.96, p_{adj} = .059$; touch vs. lift: $t(25) = 1.56, p_{adj} = .133$). However calculating the pairwise contrasts for the estimated means of the trial numbers found significant differences between the first B trial and all other trials, and between the second B trial and the second A trial (Figure 4).
Figure 4  Ratio of correctly choosing individuals in group 2. Correct choices are coded either as looking at, touching or raising the pot covering the target object. Binomial tests comparing the number of correct choices to the ratio predicted by chance (n.s.: $p_{Hoch} \geq .05$, *: $p_{Hoch} < .05$, **: $p_{Hoch} < .01$, ***: $p_{Hoch} < .001$; dashed horizontal line marks the ratio of correct choices expected by chance (50%). Generalized linear mixed model comparing the number of correct choices between trials (capital letters P, Q and R mark the trials that differ from each other significantly according to the pairwise contrasts of the estimated means, trials with the same letter do not differ significantly).

Next we analyzed whether children show an ambivalent choice behavior. For this we calculated the number of trials in which children looked at the incorrect location but subsequently touched or lifted the pot at the correct location (Figure 5). Next we compared the sum of ambivalent choices in the five trials to each other with a related-samples Cohran’s Q test. We found no significant difference when ambivalent choice was defined as looking at the incorrect location and touching the correct one ($N = 20$, $Q(4) = 2.29$, $p = .683$), however in the case when we examined the number of events when children looked at the incorrect location but lifted the pot at the correct one, we found a significant difference between the trials ($N = 20$, $Q(4) = 10.40$, $p < .05$). Post-hoc tests revealed that children showed significantly more ambivalent choice behavior in the B1 trial than in the A2 trial.
Summary

In this study we investigated whether dogs and children commit perseverative errors if they are tested in similar, simple visible displacement tasks with an identical hiding sequence to one of two locations (AABBA) and identical ostensive cues from the experimenter during each trial. We found that adult pet dogs and two year old children show a similar error pattern: their performance when the object was hidden to location A was markedly above chance, unlike the B trials. Both the dogs’ and the children’s performance dropped significantly in the first B trial and did not differ from the chance level in the second B trial.

One of the most striking results of our study was that the two year old children did not choose the correct location above the chance level in the B trials of this simple visible displacement task and that their performance in the first B trial was significantly lower than the success rate of all the other trials. This effect could not be due to a lack of a fully developed object permanence ability, since children of this age are considered to have the highest level of
this capacity (Piaget 1954). So far it has been suggested that A-not-B errors are typical in infants younger than one year of age (e.g. Sophian & Wellman, 1983).

Our results were further detailed by the analysis of ambiguous choices of children. We found that in the B1 trial there were significantly more children who looked to the incorrect location before raising the pot at the correct location than in the previous A2 trial. Therefore, two year old children’s random choices in the B trials might reflect a conflict between their more developed representational skills and the teaching aspect of the situation brought about by the ostensive cues (Topál et al. 2008).

Dogs in contrast displayed a perseverative search bias to location A in the B1 trial – this was most likely caused by location A being repeatedly enhanced earlier by the experimenter’s ostensive communication. In the B2 trial however, dogs selected randomly – possibly because by then, both A and B locations were already equally enhanced.

Even if we avoid stating that dogs are sensitive to human natural pedagogy to the same level as young children are, we still find a striking analogy between the functionally similar behavioral responses of children and dogs elicited by human ostension. Dogs are assumed to be inclined to follow social rules of the human group (Bekoff and Allen 1998; Topál et al. 2005b) and the obedient behavior of dogs and their ‘desire to please us’ is largely based on their susceptibility to human communicative signals.

Our experiment provides a new insight into how quickly episodic ostensive instructions can influence the choice preference of dogs and children alike. The results have shown that a combination of hiding order asymmetry and ostensive cues can elicit perseverative errors even in the most unambiguous situation. This finding has practical consequences for any study that employs a hide-and-search task in ostensive-communicative context with repeated hiding to a given location. One obvious solution is to refrain from communicative cues, however in many cases this is not practical (a simple and efficient attention grabber is necessary) or conflicts with the aims of the study (replacing verbal attention grabbers with other signals could be considered alien to the situation). Therefore another solution could be to take extra care in the experimental design to avoid repeated hiding to any of the potential locations during the search task, to reduce the chance that the subjects perceive any of the locations as ‘socially marked’.
STUDY 2: Image processing in a spontaneous information transfer task

Aims and questions

The principal aim of this study was to investigate whether dogs understand the referential nature of projected videos (process them in equivalence mode) and additionally to find out whether dogs can utilize information from a pre-recorded video footage to locate an object in the real world. To achieve this we employed the spontaneous information transfer test which was developed by DeLoache (1987) and extended to be used with videos by Troseth and DeLoache (1998). The test consists of a visible displacement task (Triana and Pasnak 1981) where the subject observes a demonstrator hiding an object on a video to one of multiple hiding locations. After the video demonstration, the subject is allowed to choose from the corresponding real hiding locations.

One group of dogs were tested in two conditions labeled either one-room or two-room condition. In the one-room condition the video demonstration and the hiding locations were in the same room. In the two-room condition the video demonstration and the hiding locations were in separate rooms. Dogs participated first in the one-room and then in the two-room condition. The rationale for this fixed order design was to start with the simpler one-room condition to introduce subjects to the nature of the search task with video demonstration. An additional group of dogs was tested in a control condition. The control’s setup was similar to the one-room condition, but a certain delay was introduced between the end of the video demonstration and the start of the search. The control condition was intended to control for the delay between the demonstration and the start of the search in the two-room condition, which occurred due to the dogs having to move between the two rooms.

Depending on which of the three conditions dogs are able to master several different hypotheses can be formulated about their image processing capabilities: a) Not being able to find the object reliably in any of the conditions would mean that dogs are unable to utilize information from the pre-recorded video to solve a real world task. Such a result would suggest that dogs need the projected demonstration to be interactive in order to successfully interpret the video, since they were able to follow the pointing of a human demonstrator when it was displayed via a live video connection (Pongrácz et al. 2003). b) Solving the task only in the one-room condition would mean that dogs can use the information on the pre-recorded, non-interactive video footage to find an object in the real world, but they are unable to retain this information in memory and to connect the video image with the real scene it represents (a
limited form of confusion mode). c) If dogs are able to find the object reliably in the one-room and the control conditions that would mean that they are processing the video in confusion mode. However being unable to solve the task in the two-room condition, when the video demonstration and the hiding locations are spatially separated, would show that they are not spontaneously capable of referential understanding (equivalence mode). d) If dogs are able to perform above chance level in all conditions that would support the hypothesis that similarly to humans and chimpanzees they can rely on referential understanding to solve a task.

Next to the performance of dogs we also analyzed how much attention dogs paid to the video demonstration. This was necessary because of the pre-recorded nature of videos dogs could have spent varying amount of time being attentive towards the screen, which in turn could have directly affected their success rate.

**Methods**

**Subjects**

Pet dogs (N = 36) and their owners were recruited on a voluntary basis. The dogs had to be highly motivated to retrieve a ball. Dogs were older than 1 year and represented various pure or mixed breeds (Gagnon and Doré (1992) showed that domestic dogs from various breeds showed equal performance in a visible displacement task).

Half of the subjects (N = 18, 9 females and 9 males, mean age = 3.2 years, range 1-5 years) participated first in the one-room condition and subsequently in the two-room condition. The other half of the subjects (N = 18, 9 females and 9 males, mean age = 2.8 years, range 1-6 years) participated in the control condition only.

**Setup**

All tests were performed indoors, in the experimental rooms at the Department of Ethology, Eötvös Loránd University, in Budapest. The two rooms used in this study had the same dimensions (3m×6m). In all conditions three hiding locations were used. Each one was composed of a blue plastic panel (30cm×30cm) and a plastic flower pot (diameter 12cm) which was fixed behind the panel. Each of the three panels had a different geometric shape: triangle, square or pentagon. The hiding locations were arranged along a line at a distance of 1m from each other and approximately 3m from the starting position (SP) of the dog. The position of the
individual geometric shapes was randomized for each subject in each condition. We used a small ball as the target object.

In room 1, the projector screen was placed opposite to the door: 2m wide, 1.8m high (Figure 6). Behind the screen were two loudspeakers. The projector was fixed near the ceiling on the other end of the room. In room 1 four cameras were recording the experiment. One of the cameras was an infrared camera, which recorded the dogs’ orientation during the video demonstration. An array of infrared LEDs were directed towards the dog to increase efficiency of the IR camera. In room 2 one camera directed towards the hiding locations recorded the dogs’ choices (Figure 7).

Figure 6  Arrangement of experimental room 1 in the one-room condition.
Procedure

The experiment consisted of three conditions: one-room, two-room and control. In the one-room and control conditions, both the video demonstration and the hiding locations were in room 1, in the two-room condition the video demonstration was in room 1 and the hiding locations were in room 2. Dogs participating in the one-room condition were subsequently tested in the two-room condition (after at least one week of delay). Dogs participating in the control condition were tested in that condition only. All conditions consisted of 3 warm-up and 9 test trials, each dog participated in a given condition only once.


**Warm-up phase**

The aim of this phase was to familiarize the dogs with the hide-and-search task. However our goal was to test the subjects’ spontaneous performance in the oncoming test, therefore we kept the number of warm-up trials as low as possible, to minimize the chance of any kind of learning occurring during these trials. For the same reason there was also no criterion set to pass this phase. The procedure of the warm-up phase was identical in all three conditions. In the one-room and control conditions it took place in room 1, and in the two-room condition in room 2. Each warm-up trial started with the dog, the owner (O) and the experimenter (E) being at the SP. The E showed the target object to the dog, went straight to one of the hiding locations, stopped behind it facing the SP, called the dogs attention, raised and waved the object, put it into the pot behind the hiding location and finally returned to the SP next to the O. After this the O released the dog with one command “You can go!” to search for the target object. The dog was allowed to search until the object was found. Lastly the O called the dog back, praised it and took the object from the dog. During the 3 trials the target was placed behind each hiding location once in a random order.

**Test phase**

a) One-room condition: After the warm-up trials the O covered the dog’s eyes by hand and the E turned off the lights in the room. The windows of the room were covered, therefore the room was semi-dark. The E took the object to the actual hiding location and put the object into the pot behind it, then returned to a location behind the O where he hid behind a curtain. After returning, the E started the video projection by pressing a button on the wall, and the O uncovered the eyes of the dog. The pre-recorded video was projected onto the canvas behind the hiding locations (Figure 8).

The video demonstration consisted of three phases: 1) attention getting phase (3s): the E stood still for 1s behind one of the hiding locations holding the tennis ball in his hand, then he greeted the dog saying “Hello!”, and he waved the ball saying “Look, look, look!”; 2) hiding phase (3s): the E crouched down and placed the ball to the actual hiding location, then he stood up; 3) conclusion phase (1s): the E displayed his empty hands while standing still. The E on the video placed the object to the same location where it was placed in reality. The arrangement of the hiding location shapes on the video was identical to the arrangement in the room.

After the video was over, the canvas turned black and the E turned the lights on in the room. The O released the dog and with one command allowed it search for the object. The dog
could search for the object until it was found. Then the O called the dog back and praised the dog. During the nine trials the object was placed behind each hiding location three times in a semi-random order, so that it was never at the same location in two consecutive trials.

![Image](image.png)

**Figure 8** The video demonstration from the perspective of the dog in the one-room condition.

b) Two-room condition: The three hiding locations were in room 2, but the video projection took place in room 1. After the warm-up trials in room 2, the O and the dog went to room 1, and the E stayed in room 2 and placed the target object to the actual hiding location. During this the O and the dog entered room 1 and positioned themselves in front of the door. The O ensured that the dog was facing the canvas and started the projection with a wired remote located next to the door. The video was identical to the ones used in the one-room condition, except that it showed the E in room 2 placing the target object to the actual hiding location.

After the video ended, the screen turned black and the O led the dog back to room 2. Upon entering room 2 the O released the dog, and with the one command let it search for the object. The dog was allowed to search until it found the object. Throughout the nine trials the location of the target object was randomized in the same way as in the one-room condition.
c) Control condition: The procedure was mostly identical to the one-room condition, therefore we only highlight the differences here. After the warm-up trials, the O covered the eyes of the dog and the E placed the target object to one of the hiding locations. Next the E left the room through the open door behind the O and went into an adjacent room. Upon leaving the room, the E pushed a button on the wall which started the video projection with a 5s delay. When the video started the O uncovered the eyes of the dog. After the video ended, the O and the dog left the room and took a short walk in the hallway outside of the room for the amount of time it would have taken to walk to room 2. During this time the E remotely turned on the lights of room 1. The O and the dog returned to room 1 and the O released the dog with one command to search for the target object. After the dog found the object and returned it to O, the E entered room 1 through the door behind the O, and the next trial started. Throughout the nine trials the location of the target object was randomized in the same manner used in the other two conditions.

Data collection and analysis

All trials were video recorded and the recordings coded with Solomon Coder beta (© 2012 by András Péter). Dogs’ first location choices were coded in each trial: location choice was defined as the first pot the dog looked into behind a hiding location.

We also coded whether dogs oriented towards the projector canvas during the hiding phase of the video demonstration. Orienting towards the screen was defined by the head of the dog having an angular deviation less than 45° from perpendicular to the screen. We considered the hiding phase the main section of the demonstration because in this phase the object disappeared from sight. The trials where dogs were orienting towards the screen during the entire hiding phase were labeled ‘complete attention’ trials. Trials where dogs broke eye contact with the screen were labeled ‘incomplete attention’.

Data were analyzed with IBM SPSS Statistics 21 and R 3.0.1. In each of the three conditions 4 out of the 18 videos were coded by an independent coder who was naïve regarding the aim of the study. In case of location choices there was a 100% agreement between the two coders, whereas in the case of complete / incomplete attention trials the interrater reliability was found to be: Kappa=0.82, p<0.001, 95% CI: 0.68-0.95. According to one-sample Kolmogorov–Smirnov tests with Lilliefors significance correction, the data did not follow the normal distribution, therefore we used nonparametric tests. For each condition we tested whether there
is a difference in the number of correct trials and in the number of trials with complete attention between female and male dogs (Mann-Whitney U tests). We found no difference between the sexes in any of the conditions, therefore we pooled the data for further analysis.

When comparing the number of correct trials, or the number of trials with complete attention, among the three experimental conditions, we always carried out three pairwise comparisons. The one-room and two-room conditions were compared with a related-samples Wilcoxon signed rank test, and the control condition was compared with the one-room and two-room conditions with two independent-samples Mann-Whitney U tests. We chose to analyze the data this way to account for the repeated nature of measurements in the one-room and two-room condition, and the fact that the dogs included in the control condition did not participate in either of the other two conditions. When testing for learning effects we compared the number of correct trials in the first three and the last three trials for each condition with related-samples Wilcoxon signed rank tests.

Wilcoxon and Mann-Whitney U tests require data having homogeneous variances across groups. According to the Brown-Forsythe tests both the number of correct trials (F=2.66, p=0.08) and the number of trials with complete attention (F=1.63, p=0.21) met this criteria when compared between conditions. Also the number of correct trials in the first three and last three trials met this criteria in all conditions (one-room: F=3.78, p=0.06; two-room: F=1.45, p=0.24; control: F=0.51, p=0.48). In both the Wilcoxon and Mann-Whitney U tests ties were handled in the dataset by assigning an average rank to them, and by using normal approximation.

When analyzing the effect of attention on performance, we used a generalized linear mixed model with a binary logistic link, correct/incorrect choices as the target variable, complete/incomplete attention paid to the hiding phase as the fixed effect and dog ID as a random factor and number of trial (1 to 9) set as the repeated variable. We allowed the degree of freedom to vary between tests because the differing number of trials with complete/incomplete attention resulted in an unbalanced data set. Also to compensate for potential deviations from the model’s assumptions, we used robust covariance estimates.

To account for the increased chance of type-one errors due to multiple comparisons, we adjusted the p values in each test battery using the method by Hochberg (1988) as was described by Wright (1992). The adjusted p values are marked as p_{Hoch}.
Results

First we analyzed the dogs’ performance in the three experimental conditions. We compared the number of correct trials to the level expected by chance (3 in a three choice task with nine trials) to determine whether the dogs could reliably solve the tasks (Figure 9). According to the one-sample Wilcoxon signed rank test, in the one-room condition the dogs had significantly more correct trials (N=18, Z=3.43, p_{Hoch}<0.01) than expected by chance. However in the two-room condition the number of correct trials (N=18, Z=0.31, p_{Hoch}=0.76) did not differ from chance. Finally, in the control condition the number of correct trials (N=18, Z=2.57, p_{Hoch}<0.05) was significantly higher than that expected by chance.

We also performed three pair-wise comparisons of the number of correct trials among the three conditions. When comparing the one-room and two-room conditions with a related-samples Wilcoxon signed rank test (N=18, Z=3.24, p_{Hoch}<0.01), we found that the dogs performed significantly worse in the two-room condition than in the one-room condition. After comparing the control condition with an independent-samples Mann-Whitney U test to the one-room (N=36, Z=2.22, p_{Hoch}<0.05) and the two-room conditions (N=36, Z=2.27, p_{Hoch}<0.05), we found that in the control condition the dogs had significantly more correct choices than in the two-room but significantly less than in the one-room condition.

Because of the pre-recorded nature of the video presentation, the dogs could have paid different amounts of attention to the demonstration in the three conditions, which could have caused the observed performance difference between the three conditions. Therefore we compared the number of trials with complete attention paid to the hiding phase of the demonstration between the three conditions. According to the related-samples Wilcoxon signed rank test (N=18, Z=1.78, p_{Hoch}=0.15), there was no significant difference in the number of trials with complete attention between the one-room and two-room conditions. A comparison of the control condition, with an independent-samples Mann-Whitney U test, to the one-room (N=36, Z=2.18, p_{Hoch}=0.09) and the two-room conditions (N=36, Z=1.17, p_{Hoch}=0.25), detected no significant differences between these groups either.
Number of correct trials compared to the chance level (dashed line) in the three conditions: one-sample Wilcoxon signed rank test (*: $p_{\text{Hoch}} < 0.05$; n.s.: $p_{\text{Hoch}} \geq 0.05$). Number of correct trials compared between the three conditions: related-samples Wilcoxon signed rank test (conditions differing significantly are labeled with different letters).

We also tested whether attention (orienting towards the screen) had an effect on performance in the three conditions (Figure 10) with a generalized linear mixed model. In the one-room condition the test found no differences in the number of correct trials between those with complete and incomplete attention ($N=18$, $F_{1,160}=0.26$, $p_{\text{Hoch}}=0.61$). However in the two-room condition the dogs found the object significantly more often after paying complete attention to the hiding phase of the demonstration than when they broke eye contact with the screen during this phase ($N=18$, $F_{1,119}=6.50$, $p_{\text{Hoch}}<0.05$). In the two-room condition, in the trials where they had paid complete attention to the hiding phase, the dogs had performed better (median: 33%) than when they had not watched the complete hiding phase (median 17%). In the control condition, similar to the one-room condition, we found no difference between the trials with complete and incomplete attention ($N=18$, $F_{1,80}=0.26$, $p_{\text{Hoch}}=0.68$).
Finally we analyzed whether the dogs’ performance increased during the trials in the three test conditions. We compared the number of correct trials in the first three and the last three trials for each condition. The related-samples Wilcoxon signed rank tests indicated no significant differences for the one-room (N=18, Z=0.37, \( p_{\text{Hoch}}=0.71 \)), two-room (N=18, Z=0.52, \( p_{\text{Hoch}}=0.61 \)) and control (N=18, Z=2.00, \( p_{\text{Hoch}}=0.14 \)) conditions.

**Summary**

The dogs in this study were able to find reliably the target object in the one-room and control conditions without any pre-training, except for three warm-up trials. In the two-room condition the dogs’ performance was significantly lower than in the two other conditions, although they did not orient significantly less towards the screen during the hiding phase of the demonstration. The low performance in the two-room condition cannot be attributed to the delay between the end of the video demonstration and the start of the search either, because the control condition had a similar delay, but the dogs’ performance was still significantly higher.
than in the two-room condition. This result is in accordance with previous findings showing
that dogs can reliably find a target object in a multi-well choice task with 10s or 30s of delay
(Fiset et al. 2003). Being able to find the object in the control condition means that the dogs
could memorize the physical position of the hiding location and retrieve it during search.

The result that dogs could not reliably find the object in the two-room condition
indicates that they process the videos in confusion mode according to Fagot and colleagues'
(2010) classification. This outcome is in line with the observation of Fox (1971), who found
that dogs react the same way to life sized painted dogs as to a real conspecific. However, we
also found that in the two-room condition, paying attention to the video demonstration’s hiding
phase makes it more likely that the dogs find the hidden object in the same trial. This suggests
that they memorized the relative position of the object’s disappearance on the video and
transferred this information to the relative position of the hiding locations in room 2.
Nevertheless, even in those trials where dogs paid attention to the demonstration, the median
of successful trials was only 33%. This indicates that the effect of the information transfer is
fairly small, and might only be enough to compensate for factors that would otherwise decrease
the observed performance (e.g.: choosing the location where the object was in the previous
trial).

We found no association between attention and performance in the one-room and
control conditions. This does not mean that attention affects performance differently in these
conditions than in the two-room condition, since a larger sample size could have yielded an
association in these conditions, too. On the other hand dogs might have found the hidden object
without the need to pay attention to the critical section of the demonstration, because in these
conditions the hiding locations on the video were in close proximity to the real ones.
Consequently, in this case dogs could have found the correct location by relying on simple local
enhancement cues.

These results do not support the notion unambiguously that dogs process videos only in
confusion mode. Earlier, the results of Kaminski and colleagues (2009b) suggested that dogs
can understand the referential nature of pictures. However, in that study some of the dogs
underwent considerable training before the test, which suggests that with additional training
dogs might have shown a clear sign of referential understanding in our study too. Although
currently we were interested in dogs’ spontaneous reaction to videos, it could be a topic of
future studies to find out whether training can improve dogs’ performance in a referential
understanding task.
In our present study we found that dogs could use information from pre-recorded videos, if the location of the video demonstration and the location referred by the video were in close proximity. Dogs were able to extract, memorize and retrieve location information from the video demonstration. On the other hand we did not find evidence that dogs would process life sized videos in equivalence mode, which means that to date only humans (Troseth and DeLoache 1998) and chimpanzees (Menzel et al. 1978) were shown to understand videos referentially.
STUDY 3: Image processing in a violation of expectation task

Aims and questions

In the previous study we found that dogs can reliably use information provided via an unaltered, pre-recorded, life-sized video to solve a real world search task. Our main goal was to investigate how dogs process these videos, and in order to determine this we coded dogs’ correct choices to measure their performance. With this method we found that dogs most likely process these videos in confusion mode, however similarly to Pongrácz and colleagues' (2003) study we also found indications that dogs might be able to process videos in equivalence mode.

The aim of the current study was to extend our investigation to a novel experimental paradigm to increase our understanding concerning dogs’ image processing capabilities. We planned to use the violation of expectation (VOE) paradigm this time because it enables to detect the presence of cognitive abilities which other choice based methods cannot (e.g. see Baillargeon and Graber 1988).

In a VOE test subjects are presented with scenes that show either ‘possible’ or ‘impossible’ events. Displaying different degrees of attentiveness towards the two scenes is regarded as corroboration that the subjects possess a certain type of cognitive ability (defined by the designers of the study). Therefore we continued to employ life-sized, pre-recorded videos however this time we also planned to manipulate the video sequences to produce realistic and unrealistic scenes.

There are at least two studies that that successfully used a similar method with dogs. In their study Müller and colleagues (2011) presented dogs with an object that changed its size while being occluded, and found that dogs looked longer at the object after it changed size then when it did not. However dogs were presented with real objects in this experiment and not with artificial visual stimuli. Adachi and colleagues (2007) used photographs in their VOE experiment when investigating whether dogs would connect their owners face to its voice, however this experiment involved intermodal matching of static visual stimuli with sounds (photos of human faces with voices of humans). Therefore it is still an open question how well a VOE method with using artificial and moving visual stimuli (video sequences) could be applied in cognitive experiments with dogs.

In our previous study we used choice behavior as the main response variable whereas VOE tests use looking time for similar purposes. In the previous experiment we observed dogs
on numerous occasions showing a particular behavior towards the screen that could be interpreted as surprise. This behavior could be described as orienting towards the projector canvas after the video demonstration, and was characterized by intently gazing at the projector canvas or sniffing it during the search phase. This reaction seemed to be elicited by encountering the novel experience of a life-sized video presentation and/or the unexpected appearance/disappearance of the demonstrator on the video. We observed mainly these behaviors in the early trials, suggesting that dogs habituated quickly to the novelty of the situation. Since this behavior seemed to be elicited by the surprising nature of the situation and seemed to be highly variable, we hypothesized that it could be used to assess dogs’ reaction to videos depicting expected or unexpected events.

This ‘canvas oriented’ behavior of dogs seemed to be a suitable alternative to the ‘looking time’ variable used in classical VOE tests. In the traditional VOE paradigm subjects cannot be specifically motivated to watch the presented stimuli, as any such attempt would bias their looking times. However substituting looking time with another response variable, that manifests itself after the stimulus presentation, would allow us to motivate subjects to observe the presented stimuli. For example using a similar setup as in our previous study, dogs would be motivated to watch the videos to gain information about the location of a target object, while using altered versions of a video would enable to present different scenes, possibly eliciting different canvas oriented reactions from the subjects.

Therefore we had two main goals in this study: a) to explore whether our candidate response variable could be used to gauge dogs’ responses in our VOE test; b) to explore further how dogs process projected videos. Next to the canvas oriented behavior we also planned to use correct choices as a second response variable, in order to gain a more detailed picture of dogs’ reactions towards the videos.

To address these topics we designed a task in which dogs participated in three conditions. In these the dogs were presented with three different video demonstrations showing a human hiding a target object. In the ‘realistic condition’ a human entered the room, hid the target object at one of two locations and then left the room. In the ‘upside-down condition’ dogs viewed the same video as in the ‘realistic condition’ but at this time the video was displayed upside down. In the ‘sudden appearance condition’ dogs saw a video that showed a human hiding an object to one of two locations but without showing him entering or leaving the room. In each conditions dogs had to complete six trials, all dogs participated in every condition in a counterbalanced order.
Depending on how dogs process the videos we expect them to show different reactions in the three conditions. If dog process the videos in independence mode that would mean that they only perceive them as a combination of unrelated visual elements. Therefore we would not expect dogs to orient more towards the canvas after the demonstration in any of the conditions. The number of correct choices should differ with performance being highest in those conditions where the motion cues on the video are easiest to associate with the location of the target object (sudden appearance > realistic > upside-down).

In case dogs process the videos in confusion mode, as was found earlier, then we expect them to show more canvas oriented behavior in the upside-down and sudden appearance conditions than in the realistic one. The number of correct choices should be highest in the realistic condition and lowest in the upside-down condition. The performance in the sudden appearance conditions might be influenced by the amount of ‘surprise’ (canvas oriented behavior) that dogs would show. If however dogs would perceive the videos in equivalence mode then we would expect dogs to display only slightly more canvas oriented behavior in the upside-down and sudden appearance conditions than in the realistic one. At the same time they should have the highest number of correct choices in the sudden appearance and the realistic conditions.

Furthermore there is a fourth alternative: dogs might process the three videos in different modes. This is especially likely because the upside-down videos depict a highly unusual scene. For example it can happen that dogs process the videos in the realistic and sudden appearance conditions in confusion mode but the video in the upside-down condition in independence mode. To clearly differentiate whether dogs perceive the videos in independence or confusion mode, we need a variable that can measure to which extent dogs rely on abstract visual information or on a realistically perceived scene to find the object. It is already known that retaining information in memory is more difficult with abstract than with familiar stimuli (Fei-Fei et al. 2005). Therefore if dogs perceive the upside-down videos only as a combination of abstract visual features, their ratio of correct choices should be affected by the latency between the start of the search and the salient event marking the location of the target object’s position. In both the realistic and upside-down conditions the demonstrator walks out of the scene after hiding the target object. In those cases when the demonstrator walks out on the same side where he placed the object, there is less time between the hiding and the start of dogs’ search compared to those cases when he walks out on the other side. If dogs find the object more often in those cases when the demonstrator leaves the scene on the same side where the target is, then we can
conclude that they relied on low level visual cues to solve the task (implying independence mode).

Methods

Subjects

Dogs (N=36; 18 females and 18 males) and their owners were recruited on a voluntary basis. The dogs had to be highly motivated to retrieve a ball. Dogs were older than 1 year (mean age 2.7 years, range 1–8 years) and represented various pure and mixed breeds. Prior to the experiment the owners were not informed about the aim of the study.

Setup

The tests were performed indoors, in an experimental room of the Department of Ethology, Eötvös Loránd University, Budapest (Figure 11). The room used in this study was 3m wide and 6m long. In all conditions two hiding locations were used. Each one was composed of a blue rectangular plastic panel (30cm×30cm) and a plastic flower pot (diameter 12cm) which was fixed behind the panel. The two hiding locations were 1.5m from each other and approximately 3m from the starting position of the dog. We used a small ball as the target object. The projector screen was located behind the hiding locations (2m wide, 1.8m tall) and behind the screen were two speakers. The projector was fixed near the ceiling on the other end of the room. Four cameras were recording the experiment. One of these was an infrared camera, which recorded the dogs’ orientation during the video demonstration. An array of infrared LEDs were directed towards the dog to increase efficiency of the IR camera.
Figure 11  Arrangement of the experimental room, depicting the location of the owner, dog and the experimenter at the onset of the trials.

Procedure

The experiment consisted of three conditions. All subjects participated in each experimental condition, the sequence of conditions was counterbalanced among subjects. The conditions were separated by 5 minute long breaks. The procedure of the conditions was identical, except for the projected video demonstration. All conditions were comprised of 2 warm-up trials and 6 test trials.

Warm-up trials

Each warm-up trial started with the dog, the owner (O) and the experimenter (E) being at the starting position (SP). The E showed the target object to the dog, went straight to the actual hiding locations, stopped behind it facing the SP, called the dog’s attention (“Look, look, look!”), raised and waved the object, put it into the pot behind the hiding location and finally returned to the SP next to the O. After this the O released the dog with one command (“You can go!”) to search for the target object. The dog was allowed to search until the object was found. Lastly the O called the dog back, praised it and took the object from the dog. During the 2 trials the target was placed behind both hiding location once in a random order.
Test trials

After the warm-up trials the O covered the eyes of the dog by hand, and the E turned off the lights in the room. The windows of the room were covered, therefore the room was semi-dark. The E took the object to the actual hiding location and put the object into the pot behind it, then returned to a location behind the O where he hid behind a curtain. After returning, the E started the video projection by pressing a button on the wall, and the O uncovered the eyes of the dog. The pre-recorded video was projected onto the canvas behind the hiding locations. The three conditions differed in the projected video demonstration.

1) Realistic (Rea) condition: The video started with showing the two hiding locations from the perspective of the dog on the SP. Then a human (the E) entered through one of the doors visible on the side while saying “Hello!”, walked behind the actual hiding location, turned towards the dog, raised the target object and said “Look, look, look!” Next the E on the video crouched down, placed the ball to the actual hiding location and stood up displaying his empty hands. Finally the E continued walking in the same direction as before, exited through the door on the opposite side to where he entered, and closed the door behind himself (Figure 12).

Figure 12 Six frames showing key moments of the video demonstration of the Rea condition. The frames are organized in the order of their appearance from left to right, starting in the upper left corner.
2) Upside-down (Ups) condition: in his condition the video was identical to that in the Rea condition however the video was mirrored horizontally so that the whole presentation was displayed upside-down (Figure 13). In the Rea and Ups conditions the side on which the E entered the video was alternated in the consecutive trials. The side on which he entered in the first trial was randomly chosen prior to the experiment.

![Six frames showing key moments of the video demonstration of the Ups condition. The frames are organized in the order of their appearance from left to right, starting in the upper left corner.](image)

3) Sudden appearance (App) condition: The video was similar to that in the Rea condition, however the E was not walking in or out of the room. The video started by showing the E standing right behind the actual hiding location. The E stood still for 1s then he greeted the dog saying “Hello!”, and he waved the ball saying “Look, look, look!”. Next the E crouched down, placed the ball to the actual hiding location and stood up displaying his empty hands for 1s (Figure 14).
The E on the videos placed the object to the location on the same side where it was placed in reality. After the video was over, the canvas turned black and the E turned the lights on in the room. The O released the dog without a command word and sat silent and motionless. The trial lasted until the dog returned the object to the O or until one minute has passed. At the end of the trial the E started talking to the dog which also served as a signal to the O that the trial was over. If the dog returned the object to the O, the E praised the dog together with the O. If the dog did not return the object in a minute, the E encouraged the dog to search for or return the object. During the six test trials the object was placed behind each hiding location three times in a semi-random order, so that it was never at the same location in three consecutive trials.

**Data collection and analysis**

All test sessions were video recorded and the recordings were coded with Solomon Coder beta (© 2012 by András Péter). Dogs’ first location choices were coded in each trial: location choice was defined as the first pot the dog looked into. Trials were classified as a ‘correct trial’ if the dog chose the location where the ball was. We also coded whether dogs
oriented towards the projector canvas in the search phase of the trials (after the video demonstration and before the end of the trial). Orienting towards the canvas (OTC behavior) was defined as two behaviors: a) looking into the direction of the canvas while standing still; b) sniffing at the canvas. Trials were classified as an ‘OTC trial’ if the dog showed any of the two OTC behaviors during the search phase of the trial.

The data were analyzed with IBM SPSS Statistics 21 and R 3.0.1. We analyzed the distribution of our coded variables with a generalized linear mixed model (GLMM) along three factors: a) condition (cond): what type of video demonstration was used in the session (App, Rea, Ups); b) session-sequence (s_seq): whether the session took place as the 1st, 2nd or 3rd; c) trial-sequence (t_seq): separating the first three (1-3) and last three (4-6) trials of each session. The condition and session-sequence factors had three states each and were overlapping (depending on the subject, a session could have any demonstration type and any position in the sequence) while the trial-sequence had two states and was independent from the other two factors.

Because all of the data was in binary format we used a generalized linear mixed model with a binary logistic link to test for the effect of the three factors on the number of correct and OTC trials. The target variable was the sum of the first three and the last three correct or OTC trials of each session. Condition (App, Rea, Ups), session sequence (1st, 2nd and 3rd session) and trial sequence (trials 1-3 and trials 4-6) and their interactions (Cond*S_seq, Cond*T_seq, S_seq*T_seq, Cond*S_seq*T_seq) were defined as fixed effects. To compensate for potential deviations from the model’s assumptions we used robust covariance estimates. To analyze the pattern of difference between the individual factors we contrasted the estimated marginal means produced by the model. We only calculated the pairwise contrasts for significant effects or interactions, and used the least significant difference (LSD) method to adjust for multiple comparisons (corrected p-values are marked as p_adj).

Results

First we analyzed the effect of condition, session sequence and trial sequence on the number of correct trials and the number of OTC trials (Table 2). The generalized linear mixed model indicated that the predefined fixed effects had a significant influence on all of the target variables. The number of correct trials was significantly influenced by condition, and an interaction between condition and trial sequence. The number of OTC trials was significantly
influenced by session sequence, trial sequence and an interaction of condition and session sequence.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of correct trials</th>
<th>Number of OTC trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(17,198)=3.50$</td>
<td>$F(17,198)=6.76$</td>
</tr>
<tr>
<td></td>
<td>$p&lt;0.001$</td>
<td>$p&lt;0.001$</td>
</tr>
<tr>
<td>Condition</td>
<td>$F(2,198)=6.02$</td>
<td>$F(2,198)=0.53$</td>
</tr>
<tr>
<td></td>
<td>$p&lt;0.05$</td>
<td>$p=0.59$</td>
</tr>
<tr>
<td>Session sequence</td>
<td>$F(2,198)=0.13$</td>
<td>$F(2,198)=9.25$</td>
</tr>
<tr>
<td></td>
<td>$p=0.88$</td>
<td>$p&lt;0.001$</td>
</tr>
<tr>
<td>Trial sequence</td>
<td>$F(1,198)=0.92$</td>
<td>$F(1,198)=9.96$</td>
</tr>
<tr>
<td></td>
<td>$p=0.34$</td>
<td>$p&lt;0.01$</td>
</tr>
<tr>
<td>Cond * S_seq</td>
<td>$F(4,198)=0.53$</td>
<td>$F(4,198)=3.25$</td>
</tr>
<tr>
<td></td>
<td>$p=0.72$</td>
<td>$p&lt;0.05$</td>
</tr>
<tr>
<td>Cond * T_seq</td>
<td>$F(2,198)=3.64$</td>
<td>$F(2,198)=1.26$</td>
</tr>
<tr>
<td></td>
<td>$p&lt;0.05$</td>
<td>$p=0.29$</td>
</tr>
<tr>
<td>S_seq * T_seq</td>
<td>$F(2,198)=0.02$</td>
<td>$F(2,198)=1.10$</td>
</tr>
<tr>
<td></td>
<td>$p=0.98$</td>
<td>$p=0.34$</td>
</tr>
<tr>
<td>Cond * S_seq * T_seq</td>
<td>$F(4,198)=1.88$</td>
<td>$F(4,198)=0.90$</td>
</tr>
<tr>
<td></td>
<td>$p=0.12$</td>
<td>$p=0.47$</td>
</tr>
</tbody>
</table>

Table 2 Results of the three generalized linear mixed models analyzing the effects of conditions (Cond), session sequence (S_seq), trial sequence (T_seq) and their interactions on the number of correct trials and the number of OTC. ($p$ values of significant effects and interactions are highlighted with a bold font)

As the next step we calculated the contrasts of the estimated means, for all the significant effects and interactions, starting with the number of correct trials (Table 3). When analyzing the effect of conditions, the contrasts of the estimated means showed that in the App condition performance was significantly higher than in the other two conditions. When analyzing the condition and trial sequence interaction, with trial sequence selected as the contrast field, we found that in the Ups condition dogs performed better in the early trials than in the later ones (Figure 15). When analyzing the same interaction but with conditions selected as the contrast field, we found that performance did not differ among conditions when comparing the first three trials. However when comparing the performance of the last three trials, all conditions differed
significantly from each other. In this comparison performance was the highest in the App condition and the lowest in the Ups condition.

| Condition | App. vs. Rea.: $t(198)=2.32, p_{adj}<0.05$ | App. vs. Ups.: $t(198)=3.60, p_{adj}<0.001$ | Rea. vs. Ups.: $t(198)=0.70, p_{adj}=0.48$ |
| Condition * Trial sequence (contrast field) | App. (trials 1-3 vs. 4-6): $F(1,198)=1.37, p=0.24$ | Rea. (trials 1-3 vs. 4-6): $F(1,198)=0.09, p=0.76$ | Ups. (trials 1-3 vs. 4-6): $F(1,198)=6.55, p<0.05$ |
| Condition (contrast field) * Trial sequence | Trials 1–3 $F(2,198)=0.77, p=0.46$ | App. vs. Rea. $t(198)=2.47, p_{adj}<0.05$ | App. vs. Ups. $t(198)=4.48, p_{adj}<0.001$ |
| | Trials 4–6 $F(2,198)=10.43, p<0.001$ | Rea. vs. Ups. $t(198)=2.28, p_{adj}<0.05$ |

Table 3 Results of the contrasts of estimated means for the significant effects (condition) and interactions (condition * trial sequence) of the number of correct trials. The condition and trial sequence interaction was analyzed two times, first with trial sequence selected as the contrast field, and second as condition selected as the contrast field. ($p$ values smaller than 0.05 are marked with a bold font)

We also compared the number of correct trials in the three conditions to the level expected by chance. This was performed separately for the first three and the last three trials with one-sample Wilcoxon signed rank tests and the obtained $p$-values were adjusted for multiple comparison with the Benjamini-Hochberg method (Benjamini and Hochberg 1995). The tests (App_1-3: $W=558, p_{BH}<0.01$; App_4-6: $W=589, p_{BH}<0.001$; Rea_1-3: $W=462, p_{BH}<0.05$; Rea_4-6: $W=498, p_{BH}<0.01$; Ups_1-3: $W=535, p_{BH}<0.01$; Ups_4-6: $W=324, p_{BH}=0.88$) revealed that dogs performed above chance in all conditions, except in the last three trials of the Ups condition (Figure 15).
Comparison of the number of correct trials in each condition for the first three and last three trials to that expected by chance (dashed grey line marks the chance level; * and n.s. on the bottom of the graph marks the deviation from the level expected by chance).

Next we analyzed the significant effects and interactions of the number of trials with OTC behavior (Table 4). Contrasts of the estimated means for session sequence showed that there were significantly more trials in the 1st session with OTC behavior than in the 2nd or 3rd sessions (Figure 16). Consistently with this we found that when contrasting the estimated means of trial sequence, dogs displayed more OTC behavior in the first three trials than in the last three (Figure 17). When analyzing the estimated means for the condition and session sequence interaction, with conditions set as the contrast field, we found that the conditions did not differ significantly in the 1st and 2nd sessions, however in the 3rd session the Ups condition had significantly more trials with OTC behavior than the App and Rea conditions (Figure 18). Analysis of the same interaction but with session sequence set as contrast field revealed that: in the App condition the 1st session had significantly more trials with OTC behavior than the 2nd or 3rd sessions; in the Rea condition the 3rd trial had significantly lower number of trials with
OTC behavior than the 1\textsuperscript{st} and the 2\textsuperscript{nd}; in the Ups condition the sessions did not differ from each other significantly.

<table>
<thead>
<tr>
<th>Session sequence</th>
<th>Session 1. vs. 2.: ( t(198)=3.14, p_{adj}&lt;0.01 )</th>
<th>Session 1. vs. 3.: ( t(198)=4.51, p_{adj}&lt;0.001 )</th>
<th>Session 2. vs. 3.: ( t(198)=1.84, p_{adj}=0.07 )</th>
</tr>
</thead>
</table>

| Trial sequence (trials 1-3 vs. 4-6): \( F(1,198)=9.96, p<0.01 \) |
|------------------|-----------------------------------------------|------------------------------------------------|-----------------------------------------------|

<table>
<thead>
<tr>
<th>Condition (contrast field) * Session sequence</th>
<th>Session 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F(2,198)=10,19, p&lt;0.001 )</td>
<td>( F(2,198)=1.44, p=0.24 )</td>
</tr>
<tr>
<td>Session 2</td>
<td></td>
</tr>
<tr>
<td>( F(2,198)=2.13, p=0.12 )</td>
<td></td>
</tr>
<tr>
<td>Session 3</td>
<td></td>
</tr>
<tr>
<td>( F(2,198)=3.89, p&lt;0.5 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition * Session sequence (contrast field)</th>
<th>App.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F(2,198)=14.66, p&lt;0.001 )</td>
<td>( t(198)=0.81, p_{adj}=0.42 )</td>
</tr>
<tr>
<td>Rea.</td>
<td></td>
</tr>
<tr>
<td>( F(2,198)=4.65, p&lt;0.05 )</td>
<td>( t(198)=2.76, p_{adj}&lt;0.01 )</td>
</tr>
</tbody>
</table>

| Ups. |
| \( F(2,198)=2.99, p=0.05 \)                  | \( t(198)=2.01, p_{adj}<0.05 \)               |

<table>
<thead>
<tr>
<th>Condition (contrast field) * Session sequence</th>
<th>Session 1. vs. 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t(198)=2.93, p_{adj}&lt;0.01 )</td>
<td>( t(198)=5.30, p_{adj}&lt;0.001 )</td>
</tr>
<tr>
<td>Session 2. vs. 3.</td>
<td>( t(198)=1.71, p_{adj}=0.09 )</td>
</tr>
</tbody>
</table>

| Session 1. vs. 3. | \( t(198)=2.93, p_{adj}<0.01 \)               | \( t(198)=2.58, p_{adj}<0.05 \)               |
| Session 2. vs. 3. | \( t(198)=2.37, p_{adj}<0.05 \)               |

Table 4 Results of the contrasts of estimated means for the significant effects (session sequence, trial sequence) and interactions (condition * session sequence) of the number of trials with OTC behavior. The condition and session sequence interaction was analyzed two times, first with condition selected as the contrast field, and second as session sequence selected as the contrast field. (\( p \) values smaller than 0.05 are marked with a bold font)
Figure 16  Comparison of the number of trials with OTC behavior in the three consecutive sessions regardless of experimental condition type. (capital letters A and B mark the sessions that differ significantly from each other, according to the pairwise contrasts of the estimated means, sessions sharing the same letter(s) do not differ significantly)

Figure 17  Comparison of the number of trials with OTC behavior in the first three trials (1 – 3) and last three trials (4 – 6) of all conditions. (solid horizontal line above the graph mark the results of the contrasts of estimated means; *: p<0.05;)

65
Comparison of the number of trials with OTC behavior in the three consecutive sessions between the experimental conditions. (solid horizontal lines above the graph mark the results of the contrasts of estimated means; n.s.: $p \geq 0.05$; *: $p < 0.05$; capital letters A and B mark the conditions that differ significantly from each other, according to the pairwise contrasts of the estimated means, conditions sharing the same letter(s) do not differ significantly)

After testing the possible factors affecting our two target variables we set out to analyze whether these variables influence each other. With the help of generalized linear models we tested: whether displaying OTC behavior in a trial affects performance in the same trial (Table 5). The GLMMs revealed that there was no connection between the two types of trials in either the App condition nor in the Rea condition. However in the Ups condition we found that trials with OTC behavior contained significantly lower number of correct trials than the trials without OTC behavior.
Table 5  Results of the three generalized linear mixed models analyzing the connection between the distribution of correct and OTC trials. (p-values of significant effects are highlighted with a bold font, \( p_{BH} \) marks p-values adjusted with the Benjamini-Hochberg method; the symbol ¬ marks a negative correlation)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OTC trial –</td>
<td>( F(1,214)=0.94 )</td>
<td>( F(1,214)=0.01 )</td>
<td>( F(1,214)=8.88 )</td>
</tr>
<tr>
<td>Correct trial</td>
<td>( p_{BH}=0.33 )</td>
<td>( p_{BH}=0.91 )</td>
<td>( p_{BH}&lt;0.05 ^\circ )</td>
</tr>
</tbody>
</table>

Lastly we analyzed whether the number of correct trials was influenced by the projected human’s path on the video. We separated the trials in two groups based on the video demonstration: a) trials where the projected human entered the scene on the same side where the (to be baited) target location was (identical entry); b) and trials where the human exited the scene on the same side where the (already baited) target location was (identical exit). The trials could only be separated this way in the two conditions (Rea and Ups) where the demonstrator was actually walking on the video. For each condition separately we tested whether there was a connection between the distribution of correct trials and of the trials with identical entry/exit. The two GLMMs revealed that there was no connection between the two types of trials in the Rea condition (\( F(1,214)=0.30, \ p=0.58 \)), however in the Ups condition (\( F(1,214)=9.87, \ p<0.01 \)) there were significantly more correct trials among identical exit trials.

Summary

When discussing the results of the analyses we will proceed by summarizing first the findings for the number of correct choices and afterwards the findings for the number of trials with canvas oriented behavior (OTC). Lastly we will review the results regarding the correlation of these two variables, and the effect of the demonstrator’s path on dogs’ correct choices.

The analysis of the number of correct trials revealed that dogs had the highest performance in the App (sudden appearance) condition, while their performance was lower in the Rea (realistic) and was the lowest in the Ups (upside-down) condition. However these differences were only present in the last three trials of the conditions but not in the first three. Interestingly we also found that dogs’ performance was lower in the later trials of the Ups condition than in the first ones. When comparing the number correct trials to that expected by
chance we found that dogs performed above chance in all conditions except in the last three trials of the Ups condition.

Next we continue with summarizing the findings about dogs’ OTC behavior (looking at or sniffing the canvas after the video demonstration). We found a significant decline of the frequency of this behavior both within condition and between sessions, which indicates a quick general habituation to the projected videos. Intriguingly dogs did not show a between session habituation when the Ups condition was analyzed, and corresponding to this result, in the third sessions dogs showed more OTC behavior in the Ups condition than in the other two conditions. These findings support the hypothesis that dogs did perceive that a human was displayed on the projected videos and that they reacted appropriately to the unnatural scene of the upside-down video (confusion mode). Furthermore this indicates that OTC behavior is a variable that can be reliably used to differentiate dogs’ responses to expected and unexpected scenes in a VOE task.

When analyzing whether there is a connection between the distribution of correct trials and OTC trials we found a significant association only in the Ups condition, and this correlation was negative between the two variables. Additionally we also found that in the Ups but not in the Rea condition dogs performed better in those trials when the human on the demonstration left the scene on the same side where the target object was placed. These last too findings and the earlier results (dropping performance in the Ups condition and lack of decrease of OTC behavior) all suggest that dogs treated the Ups condition very differently from the Rea and the App conditions.

When attempting to interpret these findings in the framework of image processing, we can arrive to two different conclusions. The finding that dogs performed better in the Ups (but not in the Rea) condition when the human exited on the side where the target object was, suggests that dogs processed the videos in independence mode. This is supported by the finding that a memory of a location is more susceptible to disturbances in case it is perceived only as abstract stimuli (Fei-Fei et al. 2005). On the other hand the finding that dogs show higher levels of OTC behavior in the Ups condition than in the App or Rea conditions suggests that they perceived the actual contents of the videos and reacted according to their varying level of strangeness.

There key result that can help us resolve this contradiction is that dogs’ performance dropped during the trials of the Ups condition. When encountering the first few Ups trials dogs processed the videos only as a set of independent low level visual features due to the unusual
configuration of the scene. This notion is supported by findings about the perception of biological motion with point-light figures. If a point light figure resembling a walking human is turned upside down human subjects have more difficulties identify it than when it is correctly oriented (Bertenthal and Pinto 1994; Pavlova and Sokolov 2000; Sumi 1984). The same inversion effect have been found when testing domestic chickens (Vallortigara and Regolin 2006) suggesting that it might be a universal property of animals’ visual system.

In the Ups condition while dogs were only perceiving the scene in independence mode they could still locate the target object based on the visual cues they extracted. However during the later trials dogs ‘realized’ that the videos depict a human walking upside-down (their perception switched to confusion mode processing). This kept their OTC behavior rates from dropping, and at the same time led to their performance falling as the OTC behavior interfered with their search behavior. The higher performance in the App and Rea conditions and the lack of the demonstrator’s path influencing dogs’ correct choices in the Rea condition both point towards that dogs processed the videos in those conditions in confusion or equivalence mode.

Our experiment showed that it is possible to measure dogs’ reactions towards an unexpected scene on a video via a novel method. This method was characterized by: a) providing dogs with a task which required a continuously maintained attention from them towards the videos; b) measuring their reaction to unexpected events by a set of behaviors that occurred after the video presentation, which we dubbed OTC behavior. We demonstrated that dogs react as expected to an upside-down video and found indications that their image processing might not be identical across even the same type of artificial visual stimuli, but can differ depending on the familiarity of the depicted scene.
STUDY 4: Virtual visible displacement task on the touch-screen

Aims and questions

Besides studying picture-object recognition (Aust and Huber 2006) as it was mentioned earlier, the touch-screen testing procedure has been used successfully to investigate a large number of cognitive abilities on many different animal species. This method was initially developed to be used with pigeons (Pisacreta and Rilling 1987) and since then was successfully applied to studies of many categorization and visual discrimination related abilities in these birds (for a review see Cook 2001).

A touch-screen device is comprised of three main parts: a TV screen or computer monitor, a touch sensor in front of the screen and a computer that is connected to both the sensor and the screen. In these experiments the animals are presented with various visual cues that are displayed on the screen, and the subjects are trained to touch these stimuli according to a rule set by the experimenters. Consequently the ability to learn certain rules, the speed of learning or the speed of switching from one learned rule to another, is used to detect the presence or absence of cognitive abilities or mental processes in the subjects.

The versatility of the method stems from the fact that any kind of visual stimuli can be easily displayed on the screen, allowing the usage of very large stimuli sets with ease (as compared to using photographs or slides for example). The computer allows for automated data collection from the touch sensor and stimulus presentation on the screen based on the subject’s responses. Additionally when an automatic feeder device is connected to the computer, provision of reward can be made dependent on the response of the subject, thus enabling the automation of the subject’s training.

Due to the benefits of the method the touch-screen procedure was adapted to a host of different species, though it has been mainly employed with primates and birds. During the last decades it was used to study concept learning (Bhatt and Wright 1992), numerical competence (Cantlon and Brannon 2006) and rule-guided behavior in rhesus macaques (Macaca mulatta; Buckley et al. 2009), human face perception (Tomonaga 1999), numerical competence (Biro and Matsuzawa 2001) and working memory in chimpanzees (Pan troglodytes; Inoue and Matsuzawa 2007). Among other topics it was employed to investigate the allocation of attention in blue jays (Cyanocitta cristata; Dukas 2000), visual grammar learning in keas (Nestor notabilis; Stobbe et al. 2012) and transitive inference in jackdaws (Corvus monedula; Mikolasch et al. 2013).
VIRTUAL VISIBLE DISPLACEMENT TASK ON THE TOUCH-SCREEN

In conjunction with dogs there are only two studies to date that used this method. The very first study employing the touch-screen procedure with dogs was conducted by Range and colleagues (2008). This study verified that dogs can be trained to solve tasks in this paradigm, and showed that dogs can learn to categorize natural stimuli (images of dogs vs. images of landscapes) on the touch-screen. In a following study Aust and colleagues (2008) investigated dogs’, humans’ and pigeons’ ability of inferential reasoning, and found that humans and to some extent dogs are capable of inferential reasoning by exclusion, however pigeons were not able to solve such a task.

There are two further studies that use a similar but in a few aspect different method. One study investigated whether dogs can differentiate dog faces from faces of other species (Autier-Dérian et al. 2013), however in this experiment dogs were presented with two computer screens displaying the stimuli, and while they had to touch one of the two, the touch response was recorded by a human observer. In the experiment of Nagasawa and colleagues (2011) dogs had to discriminate between smiling and neutral human faces by touching one of two pictures on the wall, however dogs were presented with photographs and their responses were not automatically recorded. Despite the differences in methodology both of these studies illustrate that using an active choice based paradigm, with artificial visual stimuli can be used successfully in experiments with dogs.

The common elements of all these studies is that dogs had to choose from two different stimuli and that they had to learn some form of discrimination of these stimuli in order to complete the tasks. However the use of computer screens allows not only for static stimuli but also for the presentation of animations, which can broaden the scope of questions that can be investigated with the touch-screen method. One area of cognitive abilities that offers itself well for testing with this method is spatial cognition and the object permanence ability. For example pigeons’ landmark learning mechanisms (Spetch 1995) and spatial search (Lechelt and Spetch 1997) was already successfully investigated this way.

Spatial cognition and the object permanence ability are most often tested with the help of visible displacement (VD) tasks. In our previous two experiments we have shown that dogs are capable to solve real world VD tasks by using information from life size projected videos. Therefore it is a logical extension of these studies to investigate whether dogs can also solve a VD task on the touch-screen. By utilizing the touch-screen method we would not only gain a novel tool to investigate spatial cognition in dogs, but it would also enable us to investigate dogs’ referential understanding further. By training dogs to solve a VD analogous task on the
touch-screen we could test whether they are able to transfer location information gained from
the touch-screen to real world locations and vice versa.

Our aim in this study was threefold. The first goal was to ascertain that dogs can learn
to solve a virtual VD task using the touch-screen device. For this we designed a task where
dogs first had to learn to touch a given stimulus on the screen, then they observed this stimulus
moving ‘behind’ one of several locations, and finally they had to choose the location where the
target stimulus disappeared.

Even if dogs could learn to solve this task, it would be possible that they use different
cognitive mechanisms (e.g. association) to do so, compared to when they searched in a real
world VD task. Therefore the second aim of this study was to test whether dogs would react to
changes of the virtual VD task in a similar manner as they would do to changes in a real VD
task. In an earlier study (Péter et al. 2016) we found that the location choice patterns of the
dogs’ were affected by the path of the human performing the hiding. When the human walked
straight to the hiding locations dogs performed equally well in all trials. However when the
human walked past, behind each location while hiding the object, dogs performed significantly
better when the target was hidden to the locations on the sides than when it was placed to one
of the locations in the middle. We planned to present dogs with similar movement patterns of
the target stimulus in the touch-screen task and to analyze if their choice patterns resembled
those found earlier in the real world tasks.

Our third question was whether dogs could select the correct location after a certain
amount of delay was introduced between the disappearance of the target and dogs’ choice. We
know from earlier studies that dogs can reliably locate a hidden object even after four minutes
of delay (Fiset et al. 2003). Being able to locate the correct target after several seconds can be
treated on one hand as further evidence that dogs solve the virtual VD task using similar
cognitive mechanisms as when they solve a real world task. On the other hand being able to
retain and recall the location of the target is also the prerequisite for further studies testing
whether dogs can utilize the location information from the touch-screen to find an object in real
space.
Methods

Experiment 1

Subjects

Adult companion dogs \((N=36)\) and their owners were recruited on voluntary basis. The criteria for the selections were: 1) the dogs had to be motivated with food and 2) the owners could come to the trainings at least once a week. The dogs were family dogs with basic obedience training and were naïve to the touch-screen testing method. From the initial 36 dogs starting the training only 13 participated in the later experiments. The other participants either canceled the training because the dogs had motivational problems or due to the owners not being able to attend the training sessions regularly \((N=15)\). Other dogs were not included because they could not reach the final stage of training during the time allocated for the experiment \((N=8)\).

The experimental group consisted of 4 males and 9 females (mean age 3.8 years, range 2–7 years), and the subjects were from 5 different recognized breeds (1 australian shepherd, 2 beagles, 4 border collies, 1 golden retriever and 1 labrador retriever) and 4 mongrels. Seven dogs had initial experience with clicker training and six did not. From these 13 subjects that took part in Experiment 1 only 10 participated fully in Experiment 2. The owners of the other 3 dogs canceled participation during the training. All dogs were maintained on a normal diet that was not changed during the testing days.

The experiment was carried out in the Clever Dog Lab research facility of the University of Vienna. Dogs were trained and tested in Experiment 1 during October 2007 and March 2008.

Apparatus

We used the same touch-screen (TS) device, which had been employed by Range and colleagues (2008). Testing was conducted in a separate room to prevent distraction of the dogs. The test apparatus consisted of a closed rectangular box housing a pellet dispenser (feeder box; 40cm wide, 70cm high, 40cm deep) and an adjacent rectangular testing enclosure (with identical dimensions), separated from the feeding box by an opaque partition (Figure 19). The testing enclosure allowed the dogs to reach the touch-screen but also shielded their vision to avoid distractions from the side and above. Inside the testing enclosure, a 15-inch TFT display was mounted onto the partition. The monitor was equipped with an infrared touch frame (Carroll Touch, Round Rock, TX; 32 vertical x 42 horizontal resolution; Huber et al. 2005;
Pisacreta and Rilling 1987). The distance between the array of light-emitting diodes and screen was 1cm. The base of the touch-screen was 42cm above the ground. For the smaller dogs there was an adjustable platform in order to reach also the top part of the screen.

Reinforcement was administered in the form of small commercial dog food pellets, which were made available through a small hole beneath the touch-screen (base was 4cm above the floor). They were delivered by the automated feeding device which was inside the feeder box. Device control was handled with a microcomputer interfaced through a digital input–output board and a custom written program (by the author).

Figure 19 Dimensions of the testing apparatus used in Experiment 1.

**Procedure**

Each subject received one or more training occasions per week. A usual training occasion took about half to one hour and consisted of three or four sessions. Each session included 29 trials (the maximum capacity of the feeding device). The training included several phases as described below. The subjects were transferred from one training phase to the next one as soon as they reliably performed on a level at, or beyond the learning criterion. The criterion required 21 or more correct first choices in 29 trials (70%) in a session.

**Operant conditioning**
Operant conditioning (with the help of the clicker method if the dog was familiar with it) and food reinforcement were used to accustom the dogs to the apparatus and to teach them to touch first the screen with their nose and then to touch the positive stimulus (S+ - a yellow star) on the screen. If the subject touched the screen a ‘positive sound’ and a piece of food followed. As there were no well-defined sessions or correct choices in this phase, dogs were transferred to the next stage of training, once they were accustomed to touch the screen with their nose.

**Precision training**

The size of the S+ was reduced step by step. The starting size was 768×768 pixels (22.5cm×22.5cm) and the final size was 250×250 pixels (7.3cm×7.3cm, Figure 20). The location of the S+ was varied randomly on the screen from trial to trial. This encouraged dogs to search visually the whole screen in order to locate the stimulus.

The sensitive area was a circle with a radius equal to the size of the S+. Touching the sensitive area resulted in a food pellet and the ‘positive sound’. When touching outside of the sensitive area the whole screen turned red and no food was received. After a correction interval of 3s followed a correction trial showing the object on the same position as before. Correction trials were repeated until the dog managed to touch the S+. Correction trials were not included in the calculation of the correct trial ratio of a session, which determined whether the dog reached the learning criterion or not.

![Example screenshots of the precision training (left), discrimination with 2 stimuli (middle) and discrimination with 3 stimuli phases (right).](image)

**Discrimination training with 2 and 3 stimuli**

In the first stage of this training 2 objects appeared on the screen (Figure 20). A positive stimulus which was always the yellow star, and a negative stimulus (S-) which was a different
geometric shape with varying color, selected randomly from a pool of shapes and colors. The size of the stimuli were 250×250 pixels, the sensitive area was a 250 pixel radius circle around the objects. Touching the sensitive area of the S+ resulted in food and the ‘positive sound’. Touching the S- resulted in no food, a red screen and a correction interval. After an erroneous choice a correction trial followed with the same objects displayed in the same positions. Correction trials were repeated until the dog touched the S+. Touching the screen outside of the sensitive areas of the stimuli resulted in no response from the apparatus. Object locations between trials were randomly generated by the software. The minimum distance between the centers of two objects was twice the diameter of the objects (500 pixels).

After reaching the learning criterion in the 2 stimuli discrimination task subjects advanced to the discrimination training with 3 objects (Figure 20). The size of the objects were reduced to 200×200 pixels (5.9cm×5.9cm) in order to have enough space on the screen for sufficient variability of object positions. The sensitive area was reduced according to the size of the objects to a 200 pixel radius circle around the objects.

**Virtual visual displacement with one location**

In this task the size of the S+ was reduced (113×113 pixels; 3.3cm×3.3cm) in order to provide sufficient space for 4 virtual hiding locations in the oncoming tasks. The task begun with displaying the S+ at the bottom center, and a white rectangular ‘hiding location’ at the upper part of the screen. The white rectangle was positioned randomly at one of four fixed locations (Figure 21). The rectangle had the same diameter as the S+. One second after it appeared on screen, the S+ started to move directly toward the rectangle and after reaching it kept on moving until it was completely covered by the rectangle (taking about 5 seconds). Food and ‘positive sound’ was provided only if the dog touched the sensitive area around the hiding location which was a 226×226 pixels large rectangle with the hiding location in the middle. Touching the screen outside the sensitive area resulted in a red screen and no food reward. After an erroneous choice a correction trial followed redisplaying the same task.

Because dogs could not be trained to wait with touching the screen while the animation was still playing, an experimenter had to hold back the dog during the movement of the S+. The experimenter sat with his back turned towards the screen and held out one arm in front of the TS apparatus, signaling to dogs that they are not allowed to touch the screen. The experimenter was unable to see the screen; therefore he relied on signal light controlled by the TS apparatus to know when the animation ended and when to allow the dog to pass.
Virtual visible displacement task on the touch-screen

One hiding location

Two hiding locations

Three hiding locations

Four hiding locations

Figure 21 Example screenshots of the ‘moving S+ with 1/2/3/4 locations’ phases showing the S+ at the beginning of the trial. Dashed red arrows indicate the path taken by the S+. During the animation the S+ moved ‘behind’ one of the visible locations until being completely ‘covered’ by it. Position of the visible hiding locations and final location of the S+ was randomized across trials.

Virtual visual displacement with 2/3/4 locations

The procedure was the same as in the previous phase except that the number of hiding locations was increased stepwise as dogs reached the learning criterion with the current number of white rectangles (Figure 21). Touching the sensitive area around the rectangle behind which the S+ disappeared resulted in food and in the ‘positive sound’. Touching the sensitive area around another rectangle results in no food, but a red screen and a correction trial. Touching the screen outside of the sensitive areas results in no response from the apparatus.

The phase with 2 hiding location was divided in two sub-phases. In the first sub-phase the two locations were never displayed in adjacent positions, there was always at least one
‘empty’ location between the two (virtual VD with two distant locations). In the second sub-phase this restriction on placing was lifted, and the two locations could be displayed at any of the four possible positions (virtual VD with two close locations). This procedure was necessary to allow dogs to familiarize with this rather difficult step of the training (having to choose between two identical looking stimuli), without the added difficulty of touching the screen with high precision.

**Test 1 – Virtual visual displacement with two different routes**

After reaching the criterion level in the previous task with 4 hiding locations, the dogs were presented with a new behavior of the S+. It did not move anymore along a straight line from the start point behind one of the hiding locations. In the first condition it moved between two locations to the upper side of the screen and then by taking a U-turn to the left or to the right it moved behind on of the adjacent locations (simple route). In the other condition (detour route) the S+ moved almost to one of the upper corners of the screen, then moved horizontally above the hiding places, and finally moved down vertically above the actual hiding place until it was completely obscured (Figure 22).

Test trials were presented in 4 trial batches. In a batch the S+ always moved according to one route type (simple or detour). The position of the target location in the first trial was either the leftmost or rightmost location. In the following three trials the position was always shifted to the adjacent position where the S+ have not been yet. Therefore in the four trials of a batch, the S+ moved behind each location once in a consecutive order. In the detour condition the S+ always started the detour from the side it disappeared in the first trial.

All dogs completed 4 batches of test trials (16 trials altogether), two batches with simple and two with detour route conditions. In one of each condition’s batches the first location was on the rightmost and in the other one on the leftmost side. During a 29 trial session dogs received two batches of test trials separated by training trials from the last training phase. Therefore the four test batches were distributed among 2 sessions. During the test trials dogs received food reward independently of which location they chose. This was done to avoid dogs’ choices in the test trials being influenced by learning about the contingencies of the task. During the intermittent training trials the rewarding scheme was identical to the one in the previous training phase. The order of the different batch types were randomized among dogs.
VIRTUAL VISIBLE DISPLACEMENT TASK ON THE TOUCH-SCREEN

Figure 22 Screenshots showing examples of the two conditions in the test trials. Red dashed lines indicate the route taken by the S+ during stimuli presentation.

Experiment 2

Subjects

This study took place one year after the end of Experiment 1. All 13 dog finishing Experiment 1 were invited to participate. The experiment was carried out in the Clever Dog Lab research facility of the University of Vienna. Experiment 2 was conducted during March 2009 and June 2009.

Apparatus

We used a TS device with different design from the one in the previous study (Figure 23). The first main change concerned the size of the screen, which was increased to a 23-inch Samsung SyncMaster NW2343 TFT screen, with an integrated EtwoTouch EIR3L23 infrared touch panel. The original TS device with the 15-inch screen was designed originally for two choice tasks only (Range et al. 2008), however in our study we consequently used four stimuli. Therefore we considered it important to enlarge the screen, for lowering the number of choice errors that were due to imprecisely executed touches. The screen was used at a resolution of 1280×800 pixels and the displayed stimuli had a size of 138×138 pixels (5.3cm×5.3cm), which meant a 150% increase in stimuli area compared to the same stimuli on the previous screen.

The second main modification a new height setting system, so that the height of the screen could be adjusted in 2cm steps, which allowed setting the screen height optimally for each dog. This made the previously used platform for smaller dogs unnecessary. Because of the adjustable screen this device had no side walls, so the auto-feeder was in a separate, closed
metal housing behind the screen separated from it by a grey plastic panel. The food pellets from the feeder were dispensed through a small hole beneath the screen. Controlling of the screen and the auto-feeder was achieved with a commercial desktop PC.

![Testing apparatus](image)

Figure 23 Picture of the testing apparatus used in Experiment 2.

**Procedure**

Each subject received one or more training occasions per week. A usual training occasion took about half to one hour and consisted of three or four sessions. Each session included 10 trials (we used shorter sessions than in Experiment 1, because the individual trials lasted longer in Experiment 2). The training included several phases as described below. Subjects were transferred from one training phase to the next one as soon as they reliably performed on a level at, or beyond the learning criterion. The criterion required 7 or more correct first choices out of 10 trials in a session (70%).

**Re-familiarization with the task**

Since the dogs did not participate for one year in touch-screen training, our first priority was to retrain them to use the new device, and to ensure that they perform on the same level as they did in Experiment 1. Therefore we presented them with the same tasks that were part of the training phase of Experiment 1 (precision training, discrimination with 2 and 3 objects,
virtual VD with 1/2/3/4 locations). Within a few training occasions (4.2±1.0 with an average of 16.4±3.0 total training sessions) all dogs reached the criterion level in the final training task (virtual VD with 4 locations).

**Virtual visual displacement with delay**

This task was identical to the virtual VD with 4 locations task in Experiment 1, with the exception that after the S+ disappeared behind the target location all hiding locations faded to black (taking: 0.5s) and after a certain delay faded back to view (taking: 0.5s). The separate stages of the training differed in the delay that was inserted between the complete disappearance and the start of the reappearance of the hiding locations. The delay ranged from 0s to 6s, and was gradually raised in 1s increments, as the dogs reached learning criterion.

Because this task proved to be unexpectedly difficult for the dogs, we had to regularly mix easier sessions between the actual ones to keep up dogs’ motivation (e.g.: present 0s delay trials, although the current training phase would have already required 2s delay).

**Data collection and analysis**

The computer that controlled the TS apparatus also recorded the stimuli chosen (touched) by the dogs in each trial. A trial was considered correct if the subject chose the S+ (or the location where it disappeared) first in that trial. Learning criterion was set to 70% or more correct trials in a session. In Experiment 1 the total number of completed sessions was calculated for each training phase. Additionally to this in Experiment 2 also the number of sessions in which dogs performed above the learning criterion was calculated. All data were tested for normality with Kolmogorov-Smirnov tests with Lilliefors significance correction. According to the results of the tests parametric or nonparametric statistical methods were used. In the test phase of Experiment 1, the number of correct trials was summed for each of the four trials, separately in the two conditions (simple route, detour route) and statistics were calculated on these values.

Data were analyzed with IBM SPSS Statistics 21 and R 3.0.1. We used a generalized linear mixed model with a binary logistic link to test for the effect of condition (simple, detour), the sequence of trials (1st, 2nd, 3rd and 4th) and their interaction (condition * trial sequence) on the number of correct trials. To analyze the pattern of difference between the individual factors we contrasted the estimated marginal means produced by the model. We only calculated the
pairwise contrasts for significant effects or interactions, and used the least significant difference (LSD) method to adjust for multiple comparisons (corrected p-values are marked as $p_{adj}$).

Results

Experiment 1

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Prec.</th>
<th>Discr. 2</th>
<th>Discr. 3</th>
<th>VVD 1</th>
<th>VVD 2d</th>
<th>VVD 2c</th>
<th>VVD 3</th>
<th>VVD 4</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bastian (m)</td>
<td>18</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>Cooper (m)</td>
<td>25</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>Foxi (f)</td>
<td>23</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td>Havanna (f)</td>
<td>30</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>58</td>
</tr>
<tr>
<td>Jamil (m)</td>
<td>21</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>Jana (f)</td>
<td>18</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>Lucy (f)</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Luke (m)</td>
<td>49</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>Luna (f)</td>
<td>24</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>51</td>
</tr>
<tr>
<td>Quinie (f)</td>
<td>30</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>Su (f)</td>
<td>31</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>Summer (f)</td>
<td>24</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>44</td>
</tr>
<tr>
<td>Toffee (f)</td>
<td>13</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>13</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>23.8±10.6</td>
<td>3.0±1.6</td>
<td>7.5±3.0</td>
<td>2.8±1.3</td>
<td>5.2±3.10</td>
<td>2.8±2.5</td>
<td>2.4±1.7</td>
<td>2.1±1.3</td>
<td>49.6±13.0</td>
</tr>
</tbody>
</table>

Table 6 Summary of the training phases. The number of sessions required by each dog to reach learning criterion (70% correct) in the training phases ((f) and (m) next to the dog’s name denote sex). Total number of sessions required by each dog to complete the entire training are displayed in the rightmost column. Mean number ($±SD$) of sessions required by dogs to complete each training phase displayed in the second row from the bottom. Results of the related-samples Friedman’s ANOVA ($N=13$, $Fr(7)=54.56$, $p<0.001$) with Dunn’s post-hoc test comparing the number of sessions in the training phases displayed in the row at the bottom: capital letters A, B and C mark the phases that differ significantly from each other, phases sharing the same letter(s) do not differ significantly.

Table 6 summarizes the results of the training phases. To advance from one training phase to the next one, dogs had to reach 70% correct or better performance in a session. First we compared the number of sessions to each other that dogs required to master each training phase. Results showed that training dogs to touch reliably a randomly placed object on the
VIRTUAL VISIBLE DISPLACEMENT TASK ON THE TOUCH-SCREEN

screen (precision training) was one of the phases that took the longest time. Also discriminating among three objects, and virtual VD with 2 distant locations required more sessions to master than most other training phases.

In the next step we tested whether the individual differences among dogs in the number of sessions required to graduate from one phase to another, could be attributed to dogs’ previous experience with clicker training, to sex or to age differences (Table 7). We did not find any effect of age or clicker training, and only an effect of sex in the virtual VD with two close locations, where male dogs needed significantly more sessions to complete the phase than female dogs.

<table>
<thead>
<tr>
<th>Age</th>
<th>Prec. *</th>
<th>Discr. 2</th>
<th>Discr. 3</th>
<th>VVD 1</th>
<th>VVD 2d</th>
<th>VVD 2c</th>
<th>VVD 3</th>
<th>VVD 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p_{\text{Hoch}} = 0.57 )</td>
<td>( p_{\text{Hoch}} = 0.32 )</td>
<td>( p_{\text{Hoch}} = 0.94 )</td>
<td>( p_{\text{Hoch}} = 0.06 )</td>
<td>( p_{\text{Hoch}} = 0.23 )</td>
<td>( p_{\text{Hoch}} = 0.32 )</td>
<td>( p_{\text{Hoch}} = 0.81 )</td>
<td>( p_{\text{Hoch}} = 0.85 )</td>
</tr>
<tr>
<td>Clicker training</td>
<td>( U = 16.0 )</td>
<td>( U = 18.5 )</td>
<td>( U = 25.5 )</td>
<td>( U = 33.0 )</td>
<td>( U = 33.0 )</td>
<td>( U = 13.0 )</td>
<td>( U = 13.5 )</td>
<td>( U = 25.0 )</td>
</tr>
<tr>
<td>( p_{\text{Hoch}} = 0.53 ) &amp; ( p_{\text{Hoch}} = 0.73 ) &amp; ( p_{\text{Hoch}} = 0.53 ) &amp; ( p_{\text{Hoch}} = 0.10 ) &amp; ( p_{\text{Hoch}} = 0.10 ) &amp; ( p_{\text{Hoch}} = 0.30 ) &amp; ( p_{\text{Hoch}} = 0.30 ) &amp; ( p_{\text{Hoch}} = 0.63 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>( U = 20.5 )</td>
<td>( U = 12.0 )</td>
<td>( U = 8.0 )</td>
<td>( U = 10.5 )</td>
<td>( U = 14.0 )</td>
<td>( U = 32.0 )</td>
<td>( U = 15.5 )</td>
<td>( U = 12.0 )</td>
</tr>
<tr>
<td>( p_{\text{Hoch}} = 0.71 ) &amp; ( p_{\text{Hoch}} = 0.41 ) &amp; ( p_{\text{Hoch}} = 0.15 ) &amp; ( p_{\text{Hoch}} = 0.26 ) &amp; ( p_{\text{Hoch}} = 0.60 ) &amp; ( p_{\text{Hoch}} &lt; 0.05 ) &amp; ( p_{\text{Hoch}} = 0.71 ) &amp; ( p_{\text{Hoch}} = 0.41 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Effect of age (Spearman correlations, \( N = 13 \)), experience with clicker training and sex (independent-samples Mann-Whitney U tests, \( N = 13 \)) on the number of sessions required to reach learning criterion (70% correct) in the training phases. \( (p \text{ values smaller than } 0.05 \text{ are marked with a bold font, } p_{\text{Hoch}} \text{ marks } p \text{ values adjusted with the method of Hochberg (1988)}) \)

Next we analyzed dogs’ performance in the test phase of the virtual VD task (Table 8). The generalized linear mixed model indicated no significant association between performance and any of the predefined factors in the virtual VD task.

<table>
<thead>
<tr>
<th>Model</th>
<th>Condition</th>
<th>Trial sequence</th>
<th>Condition * Trial seq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual VD</td>
<td>( F(7,96) = 1.36 ) &amp; ( p = 0.23 )</td>
<td>( F(1,96) = 0.37 ) &amp; ( p = 0.55 )</td>
<td>( F(3,96) = 2.28 ) &amp; ( p = 0.08 )</td>
</tr>
</tbody>
</table>

Table 8: Results of the generalized linear mixed model analyzing the effects of conditions, trial sequence and their interactions on the number of correct trials, in the virtual VD tasks.
We also compared the number of correct trials to the level expected by chance (25%) in the two conditions. According to the one-sample Wilcoxon signed rank tests dogs performed above chance in both the simple (Trial 1:  W=89.5,  \( p_{Hoch}<0.01 \); Trial 2:  W=87.5,  \( p_{Hoch}<0.01 \); Trial 3:  W=87.5,  \( p_{Hoch}<0.01 \); Trial 4:  W=91.0,  \( p_{Hoch}<0.01 \)) and the detour (Trial 1:  W= 91.0,  \( p_{Hoch}<0.01 \); Trial 2:  W=73.0,  \( p_{Hoch}<0.05 \); Trial 3:  W=79.0,  \( p_{Hoch}<0.05 \); Trial 4:  W=88.5,  \( p_{Hoch}<0.01 \)) conditions (Figure 24).

![Graph](image)

Figure 24  Dogs’ performance in the virtual VD tasks’ simple and detour conditions. Analysis was carried out separately for the two conditions. Capital letters mark the trials that differ significantly from each other, according to the pairwise contrasts of the estimated means, trials sharing the same letter(s) do not differ significantly. Dashed grey line marks the performance expected by chance; * (\( p<0.05 \)) and n.s. (\( p\geq0.05 \)) mark the results of one-sample Wilcoxon signed rank tests comparing the number of correct trials to that expected by chance.

**Experiment 2**

Table 9 summarizes the results of the training phases. At the end of the training only 2 dogs were able to reach the learning criterion (70% correct) in the virtual VD task with 6 seconds of delay (Foxi and Summer). Dogs had differing amounts of sessions during the experiment due to some owners being able to come more often to the trainings than others. However there was no significant correlation between the total number of training sessions and the longest delay at which dogs reached the learning criterion (Pearson correlation,  \( N=10 \),  \( p=0.32 \)). This result suggests the existence of individual differences in learning capacity regarding this task.
VIRTUAL VISIBLE DISPLACEMENT TASK ON THE TOUCH-SCREEN

### Summary of the virtual VD with delay experiment’s training phases.

Fractions represent the number of sessions in which dogs reached learning criterion (numerator), and the number of total sessions (denominator). Total numbers of sessions completed by each dog are displayed in the rightmost column. Number of dogs that reached a certain stage of the training by the end of the experiment is displayed in the row at the bottom.

<table>
<thead>
<tr>
<th>Dog</th>
<th>VVDd 0s</th>
<th>VVDd 1s</th>
<th>VVDd 2s</th>
<th>VVDd 3s</th>
<th>VVDd 4s</th>
<th>VVDd 5s</th>
<th>VVDd 6s</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bastian</td>
<td>17 / 20</td>
<td>9 / 12</td>
<td>4 / 8</td>
<td>0 / 1</td>
<td></td>
<td></td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>Cooper</td>
<td>6 / 19</td>
<td>0 / 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Foxi</td>
<td>8 / 10</td>
<td>12 / 13</td>
<td>4 / 4</td>
<td>4 / 5</td>
<td>2 / 3</td>
<td>1 / 1</td>
<td>1 / 1</td>
<td>37</td>
</tr>
<tr>
<td>Havanna</td>
<td>16 / 23</td>
<td>4 / 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>Jana</td>
<td>11 / 14</td>
<td>6 / 7</td>
<td>2 / 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Lucy</td>
<td>17 / 29</td>
<td>9 / 17</td>
<td>0 / 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Quinie</td>
<td>20 / 22</td>
<td>12 / 23</td>
<td>5 / 6</td>
<td>1 / 1</td>
<td></td>
<td></td>
<td></td>
<td>52</td>
</tr>
<tr>
<td>Su</td>
<td>26 / 40</td>
<td>5 / 16</td>
<td>1 / 1</td>
<td>1 / 1</td>
<td></td>
<td></td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>Summer</td>
<td>5 / 5</td>
<td>6 / 6</td>
<td>10 / 11</td>
<td>2 / 2</td>
<td>6 / 6</td>
<td>1 / 1</td>
<td>5 / 6</td>
<td>37</td>
</tr>
<tr>
<td>Toffee</td>
<td>15 / 18</td>
<td>8 / 19</td>
<td>2 / 3</td>
<td>0 / 1</td>
<td></td>
<td></td>
<td></td>
<td>41</td>
</tr>
</tbody>
</table>

| REACHED | 1      | 2      | 3      | 2      |        |

Table 9

### Summary

Considering the results of Experiment 1 we can conclude that dogs were able to learn solving a virtual VD task on the touch-screen. Detailed analysis of dogs’ learning history revealed that neither age nor prior experience with the clicker training method affected the amount of training trials dogs required to reliably solve the virtual VD task. The only difference we found was that male dogs needed significantly more sessions to complete the ‘virtual VD with 2 close locations’ phase, than female dogs. This result is in accordance with the previous findings of touch-screen experiments (Aust et al. 2008; Range et al. 2008) where females required less trials to learn the tasks, and with results of real world tasks showing that females tend to perform better (e.g. Müller et al. 2011). However it is an intriguing question why only in this one phase did we find a significant difference and not in the others, especially since this was not one of the ‘difficult’ phases as can be judged by the average number of sessions required to master it.

When comparing the number of sessions that dogs required to complete each training phase we found that the initial ‘precision training’ took the longest, followed by the ‘discrimination among three objects’ and the ‘virtual VD with 2 distant locations’ phase. None
of the virtual VD phases required more training than any of the discrimination phases. This shows that the virtual VD task which was newly introduced in our experiment was not more difficult to master for dogs than the discrimination/categorization based tasks employed in previous studies.

It is also not surprising that the ‘virtual VD with 2 distant locations’ phase was the one requiring the longest training among the virtual VD phases. This is the first phase where dogs are faced with the situation that they had to choose from two identical looking stimuli, consequently being unable to rely purely on visual perception to identify the correct target. In this phase dogs had to be able to remember the location where they have last seen the moving target, and only then could they reliably select the correct stimulus. When increasing the number of potential locations to three and four dogs actually required the least number of training trials compared to other phases. This shows that once dogs got familiar with the logic of the task in the two location virtual VD phase, they required almost no effort to reliably choose the correct location in a task with a larger number of potential locations.

By analyzing the results of the test trials with modified route (Simple and Detour condition) we found no difference between the two conditions or between the trials of the individual conditions. When comparing the rate of correct choices to the level expected by chance, we found that dogs selected the correct location above chance in every trial of the conditions. These findings are in contrast with those obtained in the real world equivalent of the task (Péter et al. 2016). In the real world VD task we found that dogs performed significantly worse in the Detour condition than in the Simple condition. Although in the Detour condition of virtual VD task in the 2nd and 3rd trials dogs performed seemingly worse than in the 1st and 4th, which is a pattern similar to what we found in the real counterpart of the task, however the observed differences between trials were statistically not significant at the current sample size.

The reason behind this difference could be attributed to many potential factors. Firstly the task on the touch-screen was asocial and we know that dogs’ choices are more easily influenced in a social context (Topál et al. 2009a). Secondly in the real world version of the task the human carrying the target object returned to the starting position before the dog was released to search. Maybe including a carrying agent (a second geometrical shape that moved together with the target stimulus but would have returned to the starting position after the target disappeared behind one of the locations) would have produced similar patterns as observed in the real world task. Thirdly dogs had extensive training with the task in the virtual setup but no or minimal prior exposition to a VD tasks in the real life setup. This could have boosted dogs’
performance in the virtual VD tests and masked the effect the change of route had on the number of their correct choices.

Experiment 2 was conducted after one year of the end of Experiment 1. Because of this hiatus we included a retraining phase before Experiment 2 to refamiliarize dogs with the touch-screen procedure. The retraining phase lasted considerable shorter than the original training phase. This finding suggests that dogs retained most of what they have learned during the original training, and means that once a dog acquires the necessary skills to work with the touch-screen device, it will be able to participate in newer touch-screen studies without the need to repeat the work intensive initial training. On the other hand this can also serve as a warning, showing that what dogs learned during a touch-screen task could potentially influence the outcome of any later touch-screen tests. At the same time from our retraining data we are not able to discern which parts of the touch-screen procedure did the dogs retain precisely (e.g. only the general skill of how to interact with the device or more task specific details).

The most surprising result of Experiment 2 was that how difficult it was for dogs to locate the position of the target even after a few seconds of delay. Only two of the ten subjects were able to find reliably the correct location after a 6s delay. This finding is even more striking compared to the 240s delay after which dogs could reliably find a hidden object in a real world VD tasks (Fiset et al. 2003). One possible explanation for this finding is that because dogs were exposed to a large number of virtual VD trials without delay, they adopted a strategy where they used low level cues (location where the target was last seen moving) and orienting their heads towards this location (as was found in early studies: Hunter 1913; Walton 1915) to choose the correct stimulus. However in the condition where the four locations became invisible they would have needed to change their strategy, and rely on memorizing the target location. This theory could explain dogs having initial difficulties with the delayed virtual VD task, however according to our results, dogs were never able to make this switch. An observation of those two dogs that performed best in the task, also gave the impression that they choose the correct location only when they kept their heads motionlessly oriented towards the screen.

We found no correlation between the number of total training sessions and the highest amount of delay at which dogs could reliably choose the correct location. On the one hand this means that, albeit dogs had varying number of training trials in this experiment, the fact that only so few could solve the task with longer delays is not attributable to the lack of training. On the other hand it also implies that giving dogs the opportunity to participate in even more training would have probably not increased their performance considerably.
GENERAL DISCUSSION

The aim of our first study was to investigate whether two year old children and dogs would commit perseverative errors in an ostensive-communicative context in a simple visible displacement task with two locations and a fixed hiding sequence (AABBA). We found that both children and dogs showed a significantly lower performance in their first B trial and their performance did not differ from chance in the second B trial.

It was shown earlier that the lack of ostensive cues during a test decreases the tendency to choose location A in a B trial both in dogs and in 10 month old babies (Topál et al. 2009a). However, the asymmetry between providing ostensive cues only when hiding to location A and not when hiding to B, made it difficult to differentiate between the possible effect of ostensive communication itself and the interaction of ostensive cues and spatial asymmetries (such as local enhancement – Marshall-Pescini et al., 2010). However our findings support strongly the notion that ostensive communication plays a significant role in the emergence of A-not-B error in both dogs and children (Kis et al. 2012; Topál et al. 2010; Topál et al. 2009c; Topál et al. 2009a; Topál et al. 2008). Whether or not this effect is mediated by the same or homologous cognitive processes in children and dogs remains unclear. A parsimonious explanation could be that while the two year old children’s reaction was caused by the natural pedagogy aspect of human ostension, dogs were reacting to an ‘episodic instruction’ of the human demonstrator.

Our results have demonstrated that a combination of ostensive cues and hiding order asymmetry alone can elicit perseverative errors even in a simple visible displacement task. Therefore our findings have practical implications for studies that use a similar setup. Experimenters should either omit ostensive cues or if this is unfeasible they should refrain from placing the target object repeatedly to the same location. In accordance with these recommendations in both of Study 2 and 3 we employed a hiding order where the target was not placed to the same location twice in a row.

In the second and third study we investigated how dogs perceive life-sized projected videos. For the classification of how dogs are processing images we used the definitions of Fagot and colleagues (2010) who differentiated between three levels of image processing: independence, confusion and equivalence. In independence mode an animal perceives an image only as a combination of features and patterns, independently of what the image might depict. In confusion mode the animal perceives the image as being identical to the object it depicts. In
equivalence mode the animal comprehends that the image is a representation of the depicted object (this mode is also called referential understanding).

In Study 2 we used a spontaneous response task, based on the design of Troseth and DeLoache’s (1998) study. In our task dogs have seen a video of a human demonstrator hiding an object in the same (one-room condition) or in a different room (two-room condition), and then were prompted to search for the object in the room corresponding to the video. Finding the object in the two-room condition, would have demonstrated that dogs are spontaneously able to process images in equivalence mode and thus to be capable of referential understanding. In a control condition, that was similar to the one-room condition, we inserted a delay between the end of the demonstration and the start of the search. This condition served to ascertain that dogs are capable to store, retain and recall from memory the information about the location of the object.

We found that dogs could reliably find the object in the one-room and the control, but not in the two-room condition. This result does not support the notion that dogs could spontaneously process images in equivalence mode. At the same time we could also exclude that dogs process the images in independence mode. Perceiving the videos only as unrelated visual cues would have required dogs to associate these cues with the location of the object throughout the trials, which in turn would have manifested as a performance increase during trials. However we found no indication of increasing performance in any of the conditions.

By excluding independence mode and finding no conclusive proof of equivalence mode we could conclude that dogs processed the videos in confusion mode. However we also found that in the two-room condition the likelihood of finding the object depended on how much attention dogs paid to the video demonstration beforehand. This finding suggests that dogs could in some way still connect the video seen in room one to the position of the hiding locations in room two. This finding could be treated as an indication that under different conditions dogs might be able to comprehend the connection between the projected video and the particular location it depicts.

In Study 3 we aimed to investigate the question of image processing in dogs from a different perspective. In this study we employed a violation of expectation task where dogs were presented with three types of life-sized, projected videos: realistic, upside-down and sudden appearance. They were motivated to watch the videos by an object search task similar to the one-room condition in the previous study. We recorded dogs’ correct choices in the object
search task and their orientation reactions towards the projector canvas after the demonstration. According to the logic of the VOE tasks we expected dogs showing more orientation behavior to the unexpected than to the expected scenes (Baillargeon et al. 1985). Therefore if dogs recognized the real scene depicted in the upside-down video (confusion or equivalence mode) and not only a set of unrelated visual stimuli (independence mode) than we have expected dogs to show more orientation towards the upside-down video than to the other two types.

According to our predictions dogs oriented more towards the upside-down video, however only if the video was presented last in the sequence of the three presentation types. We also found indications that dogs processed the video in this condition in independence mode. Considering the results we concluded that dogs most likely processed the videos in the early trials of the upside-down condition in independence mode and only in the later trials did they realize what the video depicted (switched to confusion mode processing).

Therefore the most surprising finding of this experiment was that dogs can process the same type of artificial visual stimulus (life sized videos) in different modes depending on the familiarity of the depicted scene (confusion mode in the realistic and sudden appearance conditions and independence mode in the early trials of the upside-down condition). It is also likely that their processing mode can change once they have the chance to encounter a certain stimulus multiple times (changing from independence to confusion mode during the trials of the upside-down condition).

Summarizing the findings of Study 2 and 3 we can conclude that dogs primarily process life sized projected videos in confusion mode. This outcome is in line with the observation of Fox (1971), who found that dogs react the same way to life sized painted dogs as to a real conspecific. However we also found that if the depicted scene is alien to the situation then dogs might ‘revert’ to independence mode processing, but switch to confusion mode once they get more familiar with the stimulus. This later finding suggests that image processing of dogs can depend on the familiarity with the stimulus, the same way as humans who have never seen pictures have initial difficulties recognizing their content (Miller 1973).

We were also able to demonstrate the feasibility of the spontaneous information transfer task developed by Troseth and DeLoache (1998) to test image processing in animals. This task was only employed with humans, orangutans and chimpanzees to date (Poss and Rochat 2003), and ours is the first study that extended this method to a non-primate species. This shows that
the spontaneous information transfer task could potentially be used to test the image processing and referential capabilities of many other species.

By utilizing ‘canvas oriented behavior’ instead of ‘looking time’ as a response variable in our third study, we pioneered a new method to conduct violation of expectation experiments with dogs. The novelty of our method lies in using an object search task to direct dogs’ attention towards the presented videos (instead of other attention grabbers) and using a behavior that occurs after the video presentation ended as a measure of how unexpected the stimuli were perceived by dogs.

In the two experiments of Study 4 we investigated whether dogs could learn to solve a virtual displacement analogue task using a touch-screen device. This was prompted by our findings in the previous studies which showed that dogs were able to utilize information from artificial visual stimuli for solving a real world visual displacement task. On the one hand we aimed to extend the scope of the touch-screen studies with dogs from classification based questions utilizing static images (e.g. Range et al. 2008) to different topics that could be addressed with the help of computer generated animations (e.g. maze solving: Miyata et al. 2006; tracking of a moving object: Wilkinson and Kirkpatrick 2009). We planned to compare the performance of dogs in different versions of a virtual visible displacement task, with already known performance data from analogous real world tasks, to determine whether dogs solved the virtual task using the same cognitive mechanisms as they did during the real tasks. On the other hand once dogs would have learned to solve the task on the touch screen we hoped to be able to investigate whether dogs could solve a real world visual displacement task by using information gained from the touch-screen. This method would have allowed us to investigate the referential understanding in dogs with a type of artificial visual stimuli differing from the one utilized earlier.

In Experiment 1 we concluded that dogs were successful in learning to solve the virtual visible displacement task. Additionally we found that it was not more difficult to master this task than to learn a simple discrimination, based on the number of sessions dogs required to reach learning criterion. Next we analyzed dogs’ performance in a modified versions of the task where the target stimulus moved on a new path. From earlier visible displacement tasks (Péter et al. 2016) we knew that depending on the path of the person who was hiding the object, dogs show different search patterns. However, in the touch-screen experiment we found no differences in the number of dogs’ correct choices depending on the target stimulus’ path in the virtual task.
Experiment 2 was conducted one year after the first one. This gave us the opportunity to compare how quickly the dogs can be re-trained, to reach their original performance level compared to how long it took to train them originally to perform over the criterion level. We found that dogs needed three times less sessions for the re-learning, than acquiring newly the task, suggesting that they retained most of the skills they had obtained during their original training one year ago. However our main goal in this experiment was to test whether dogs would be able retain and recall the memory of the correct location, even after a certain delay was included between the disappearance of the target stimulus and the start of dogs’ search. We know that in a real life search tasks dogs are capable of locating a hidden object even after four minutes of delay (Fiset et al. 2003). Surprisingly we found that in our virtual visible displacement task dogs had serious difficulties choosing the correct location even after a few seconds of delay. The longest delay at which two of our subjects could solve the task was a mere six seconds.

By combining the results that (1) dogs have difficulties solving the virtual search task after even a short delay, and (2) their performance is not affected by the path of the target stimuli, we can conclude that they solve this task differently from a real world task. While in a real world visible displacement tasks dogs use their object permanence ability, in the virtual version of the task they most likely use low level visual cues (location where the target was last seen moving) and keeping their heads oriented towards this location (Hunter 1913; Walton 1915) to choose the correct location.

These results suggests that the touch-screen method might not be an ideal solution to test physical cognition in dogs. This could be largely caused by the touch-screen procedure, which is encouraging the use of solutions that rely on the visual cues available for immediate perception and further reinforcing these strategies during the extended training sessions. It might be possible to teach subjects to change these simple mental strategies for more complex ones, however in our study we were not able to achieve this. Still the touch-screen method remains a successful paradigm for testing abilities related to visual categorization, as has been shown by studies on dogs and numerous other species (e.g. Aust et al. 2008; Cook 2001).

In the next section we will discuss some of the questions that were raised by our findings and that could not be related to the hypotheses of the studies. In Study 1 we found that children and dogs alike commit perseverative errors in an ostensive-communicative context in a simple visible displacement task. According to the ‘natural pedagogy’ reasoning of Topál and colleagues (2008) we argued that children commit these errors because in a pedagogical context
they are quick to extract and follow social rules. However it is not reasonable to assume that dogs are guided by the same cognitive mechanism as infants in this test. What is clear that dogs only commit these errors in a communicative context (Kis et al. 2012) but how they interpret the humans ostensive cues is unclear. In their experiment Scheider and colleagues (2011) asked the same question about why dogs follow the human pointing gesture. They hypothesized that dogs either perceive the human pointing as an imperative or as an informative gesture and their results seemed to support the later notion. We can ask the same question about why dogs are inclined to follow a rule of choosing the ‘socially marked’ A locations in the B trials. A study based on the methodology of Scheider and colleagues (2011) varying the presence of the target object and the tone of verbal utterances (more or less imperative) would have the potential to answer these questions.

In Study 2 we found indications that dogs were able to use information gained from a video demonstration in one room to find an object in another room. This finding did not provide enough evidence to conclude that dogs used referential understanding, but neither could we exclude the possibility that dogs would show evidence of this ability under different conditions. Simultaneously, in Study 3, we concluded that exposure to videos can lead to a change in how dogs process these stimuli. We found that dogs switched from independence mode to confusion mode processing, likewise it is also possible that similarly to pictorially naïve humans (Miller 1973) dogs could also switch from confusion mode to equivalence mode processing through exposure. This possibility is also supported by the results of Kaminski and colleagues (2009b), who have shown that dogs are able to recognize the correspondence between photographs and the objects they depict. Therefore it is not unlikely that a more intensive initial exposure to videos, showing events in a different room, could lead to dogs connecting the videos with the remote location they depict, thus displaying an ability for referential understanding.

Another possibility to bring to light any latent ability for referential understanding is to introduce response contingencies. In the task we used in Study 2 dogs were allowed to search for the object until they found it. This means that, since finding the object was the reward, dogs were positively reinforced in each trial regardless of their performance. Consequently this could have led to dogs not using cognitively demanding mechanisms to find the target, but to resort to a random or a rigid search pattern. Kuhlmeier and colleagues (1999) found in an experiment similar to ours that chimpanzees searching for a reward resorted to fixed search patterns instead of using information provided on a scale model. However when modifying the task in a way that subjects only received a reward when selecting the correct location first, some of the
previously unsuccessful chimpanzees started to use the information from the scale model (Kuhlmeier and Boysen 2001). It would be a promising direction for future studies to modify our spontaneous information transfer task to included response contingencies, and to test whether dogs would exhibit referential understanding under such conditions.

A third possibility is to introduce an intermediate step between our two conditions where the projected video was behind the real locations (one-room) and where the real locations were in a different room (two-room). It has been shown that dogs are able to locate a hidden object in a visible displacement task if the hiding locations are rotated by 90° (Miller et al. 2009a), additionally dogs could reliably locate the object even after 15s of delay in this task (Miller et al. 2009c). It would be possible to modify our one-room condition by showing dogs a video of a human hiding the target object which was recorded from a perspective rotated by 90° relative to the dog’s perspective of the hiding locations. This way the physical position of the demonstrator on the video and of the hiding locations would be disassociated. Such a test could show if dogs are able to transpose the spatial information from the video to the physical locations in reality and could also serve as an additional training step between the more simple one-room and more complex two-room conditions.

In Study 3 when analyzing dogs’ canvas oriented (surprise) behavior we found that they exhibit this reaction more often in the condition with the upside down video, however this difference was only significant in the last session. One possible explanation for this phenomenon is that dogs required exposure to normally oriented videos to form an expectation about what they are going to see. Therefore only after seeing such videos in two sessions did dogs form the general knowledge that they are going to see a ‘correct-end up’ human. This explanation, however, is unlikely because throughout their lives dogs have only seen humans walking this way, therefore an upside down walking human should have been an unexpected stimulus no matter when they encountered it.

Another possible explanation is that since the dogs were naïve to projected videos, they displayed a high rate of canvas oriented behavior initially due to the novelty of the situation, regardless of the content of the video image. These high levels of indiscriminative reactions could have masked any canvas oriented behavior that would have been elicited by the unexpectedness of the content of the videos. One way to decide between these two possibilities is to design an experiment where dogs are either confronted with two sessions of upside down videos followed by one session of correct-end up videos, or similarly to Study 3 they first observe two sessions of correct-end up and then one session of upside down videos. If dogs
show higher levels of canvas oriented behavior in the later condition, then we can conclude that they show these reactions because it is unexpected compared to their real life experiences (as opposed to being unexpected compared to the previously seen videos). This in turn would give further support to the notion that dogs are able to connect the scenes displayed in videos with real life events.

In the 2nd Experiment of Study 4 we found that dogs are not able to select on the touch-screen the location marked by a moving target in a virtual visible displacement task. We explained this outcome with dogs using only low level visual cues (following the motion of the target stimulus) and fixating their heads in the direction of the correct location during the delay. We argued that during the training phases without delay, dogs could successfully solve the task this way and were unable to switch to a different cognitive strategy (e.g. using their object permanence ability) to solve the delayed task. This explanation, however, still leaves the question open why dogs were unable to make this switch. This question is even more intriguing considering that in Study 2, where dogs could observe a video of a human hiding an object, they still were able to choose the correct location after multiple seconds of delay. There are two major differences between these studies which could be responsible for the differences observed in dogs’ performance.

The first of these differences is that the video demonstration in Study 2 was more detailed and more life-like than the simple animation in Study 4. These details were that a human was carrying the target object, in the background the experimental room was visible, the video provided monocular depth cues, etc. The additional visual cues on the video could have helped dogs relate the scene to real world experiences they had with hide and search tasks, with the experimental room itself or with the hiding locations.

The second important difference is that in Study 2 the hiding locations which dogs had to choose were real, whereas in Study 4 dogs had to choose from virtual locations. The virtual locations (white rectangles) in Study 4 lacked any resemblance to a real object or any visual cues that would have enabled them being perceived as 3D objects. Therefore it is possible that dogs had difficulties to represent the white rectangles as objects in space, which in turn would have restricted them from utilizing their object permanence or spatial cognitive abilities to remember the correct location.

This phenomenon is partly similar to the ‘video deficit effect’ observed in young (typically two year old) children. It was found that children have more difficulties with finding
an object when they observe the hiding through television (Troseth and DeLoache 1998) and that they are able to imitate an action to a lesser extent (Hayne et al. 2003; Schmitt and Anderson 2002), than when they observe the demonstrations in real life or through a window. Explanations of this phenomenon either state that children do not perceive the images in a television as real (Troseth and DeLoache 1998) or that they perceive the objects and events on television as being confined to the image (Jaglom and Gardner 1981). It is therefore possible that dogs perceive stimuli on smaller screens (television, computer monitor) in a similar manner as two year old children. The finding that dogs failed to show contagious yawning when observing humans yawn on a laptop screen (Harr et al. 2009) gives additional support to this notion.

SUMMARY

In the last chapter we will review the most important findings and conclusions of this thesis. In our first study we have shown that in the presence of ostensive communicative cues two year old children and dogs commit perseverative errors even in the most symmetric and unambiguous experimental arrangement. With this finding we provided further support for children being influenced by natural pedagogy (Csibra and Gergely 2009) in social communicative contexts and dogs being extremely sensitive to regularities of a task under similar conditions. These findings at the same time underscore the importance of controlled stimulus presentation, showing how sensitive subjects can be to details of a task.

In the second and third studies we investigated how dogs perceive life-sized projected videos. We found that life-like videos are processed in confusion mode, and while unfamiliar scenes are initially processed in independence mode, exposure can change the processing of these videos. Furthermore we also wanted to find out whether dogs would be spontaneously capable of referential understanding, however we did not find conclusive proof of this ability. Since we did not try every conceivable approach to determine whether dogs are able to show referential understanding (e.g. increased exposure to videos, introducing response contingencies) further studies are required to settle this question adequately.

From a methodological standpoint by utilizing successfully the ‘spontaneous information transfer task’ with dogs we demonstrated that it can form a solid basis for studying image processing in animals. Hopefully this finding will prompt researchers to use this task with more species in the future. Also by showing an alternative method to conduct violation of
expectation tests with dogs in our third study, we hope to provide a new tool for the investigation of dogs’ cognitive capacities. A promising candidate of such experiments would be the study of physical cognition (e.g. causality or spatial cognition), where dogs could be presented with the human on the video interacting with objects disobeying these rules, and dogs’ reactions to such scenes could be recorded.

In our last study we have shown that dogs could learn to solve a visual displacement analogue task on the touch-screen, however we found that dogs did not utilize the same cognitive mechanism to solve these tasks than real world visual displacement tasks. We concluded that the most likely reason for this lies in the extensive pre-training required to acquire touch-screen tasks and in the overly abstract nature of the presented stimuli.

The finding of the second and third study showing that dogs process life-like artificial stimuli in confusion mode supports the already prevalent practice to use images as a replacement for real stimuli. In studies that use life-sized images researchers can confidently expect that dogs will react to artificial stimuli comparably as to its real counterpart (e.g. Faragó et al. 2010; Pongrácz et al. 2003). However, by considering the findings of our touch-screen study and the varying success of experiments with smaller than life-sized images (Harr et al. 2009; Téglás et al. 2012) more research is required to determine that under which conditions do dogs perceive small images as corresponding to their depicted content.

Our first study demonstrated that just how easily subjects can be influenced in a communicative context. The cumulative findings of our second, third and fourth studies have shown that with proper consideration artificial visual stimuli can be used as a replacement for real stimuli, providing all the benefits associated with this method (repeatability, stimulus control). Therefore next to the previously mentioned studies of physical cognition, also studies of social cognition could profit from this methodology, because the non-interactive nature of the stimulus presentation could eliminate many sources of the Clever Hans effect (Pfungst 1911). Additionally it is known that the precise timing of ostensive cues can have a dramatic effect on dogs performance (Range et al. 2009) and using video demonstrations provides the means to have the timing of cues under perfect control.

Another method that was not employed with dogs before, but which has great potential in studies of social cognition and communication, is the interactive playback method (Ord and Evans 2002). Instead of being presented with an invariable video sequence, subjects would see different pre-recorded videos depending on their behavior. Such a method could combine the
benefits of artificial visual stimuli with the responsiveness of a real social partner. Since the technology to design such experiments is readily available, there is nothing standing in the way of experiments pioneering this method with dogs.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Péter Pongrácz for the professional input he provided throughout the conducting of the experiments and the writing of this thesis. I would like to thank my family for their support that enabled me to pursue my career as a researcher. I would like to thank my loving girlfriend Lisa Horn, for standing beside me and believing in me even when I lost courage. I would like to thank my colleagues at the Department of Ethology in Budapest and at the University of Vienna for helping with the experiments and for their comments and ideas. And finally I would like to thank all the dog owners and parents who devoted their time to participate in the studies.

The studies of this theses were funded by a grant from the Hungarian Ministry of Education OTKA K82020.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Horn L, Range F, Huber L (2013) Dogs’ attention towards humans depends on their relationship, not only on social familiarity. Animal Cognition. DOI: 10.1007/s10071-012-0584-9


Hunter WS (1913) The delayed reaction in animals and children. DOI: 10.5962/bhl.title.45847
REFERENCES


Kaminski J, Bräuer J, Call J, Tomasello M (2009a) Domestic dogs are sensitive to a human’s perspective. Behaviour. DOI: 10.1163/156853908X395530


Kubinyi E, Topál J, Miklósi Á, Csányi V (2003b) Dogs (Canis familiaris) learn their owners via observation in a manipulation task. *Journal of Comparative Psychology*. DOI: 10.1037/0735-7036.117.2.156


REFERENCES


Menzel EW, Premack D, Woodruff G (1978) Map Reading by Chimpanzees. Folia Primatologica. DOI: 10.1159/000155845


107


Miller RJ (1973) Cross-cultural research in the perception of pictorial materials. Psychological Bulletin. DOI: 10.1037/h0034739


REFERENCES


Neuhaus W (1953) Über die Riechschärfe des Hundes für Fettsäuren. Zeitschrift für vergleichende Physiologie. DOI: 10.1007/BF00350805


Péter A, Topál J, Miklósi Á, Pongrácz P (2016) I saw where you have been—The topography of human demonstration affects dogs’ search patterns and perseverative errors. Behavioural Processes. DOI: 10.1016/j.beproc.2016.02.005


Pfungst O (1911) Clever Hans (the horse of Mr. Von Osten) a contribution to experimental animal and human psychology. DOI: 10.5962/bhl.title.56164

Piaget J (1954) The construction of reality in the child. DOI: 10.1037/11168-000

Pilley JW, Reid AK (2011) Border collie comprehends object names as verbal referents. Behavioural Processes. DOI: 10.1016/j.beproc.2010.11.007
REFERENCES


REFERENCES


REFERENCES


REFERENCES


ABSTRACT

This thesis presents the results of four studies: three focus on how dogs (*Canis familiaris*) perceive and process artificial visual stimuli, and one investigates how dogs and children are influenced by ostensive communicative cues in an object search task.

Study 1 focused on the question whether adult dogs and two-year old children commit perseverative errors in a two way choice task where visual and auditory cues are presented in a symmetrical and unambiguous way (in contrast to previous studies e.g. Topál et al. 2008). The results showed that both dogs and children commit perseverative errors in such a task, and demonstrated how easily subjects’ performance can be influenced in an ostensive context.

The starting point of Study 2 and Study 3 was the fact that to date, several experiments with dogs utilize images to replace real stimuli; however no investigation was carried out to determine how dogs perceive and process these artificial stimuli. By using the classification of Fagot and colleagues (2010) we aimed to find the answer to this question.

In Study 2 dogs observed a human hiding an object on a life-sized projected video, and were allowed to search for the object in the room depicted on the screen. Dogs reliably found the object when the video demonstration and the hiding locations were in the same room, but did not when these were in different rooms. From the results we concluded that dogs processed the videos in confusion mode (treated the projected video as the same as the content it depicted), and did not comprehend the referential nature of the stimuli.

Study 3 examined whether in an object search task, similar to the one employed in Study 2, dogs would react differently to videos showing a human upside down or correct-end up. By analyzing dogs’ orientation towards the projector canvas, and dogs’ performance, we found that in the upside down condition dogs initially processed the videos in independence mode (perceiving the images only as set of unrelated visual cues) and only after some exposure to the videos did dogs start to process them in confusion mode.

In Study 4 we conducted virtual object search tasks with dogs, using the touch-screen testing procedure. After comparing dogs’ performance in multiple versions of the virtual visible displacement task with results of similar real world tests, we concluded that dogs used different cognitive mechanisms to solve the tasks on the touch-screen as opposed to solving them in real life.
ÖSSZEFOGLALÓ

Ebben az értekezésben négy kutatás eredményeit mutatjuk be: három a kutyák képérzékelését és feldolgozását vizsgálja, egy pedig arra keresi a választ, hogy milyen módon hat gyermekek és kutyák választási viselkedésére egy osztenzív kommunikatív helyzet.

Az első vizsgálatban azt elemeztük, hogy két éves gyermekek és egy évnél idősebb kutyák követnek-e el perszeveratív hibákat egy olyan tárgykeresési teszben amelyet, az eddigi kutatásoknál (pl. Topál et al. 2008), szimmetrikusabb és egyértelműbb inger prezentáció jellemez. Azt találtuk, hogy a kutyák és a gyermekek is elkövetnek perszeveratív hibákat ebben az egyszerűsített helyzetben, és egyben rávilágítottunk arra, hogy milyen könnyen befolyásolható az alanyok teljesítménye egy ilyen tárgykeresési teszten.

A második és a harmadik vizsgálat kiindulópontját az adta, hogy bár már számos olyan vizsgálatot végeztek kutyákkal amelyekben mesterséges vizuális ingereket használtak valódi ingerek helyettesítésére, azonban még egy kutatás sem vizsgálta, hogy a kutyák miképpen érzékelik és dolgozzák fel ezeket a képeket.

A második vizsgálatban a kutyák egy embert láthattak egy életnagyságú videón amint elrejt egy tárgyat, majd megkereshtették ezt a tárgyat a videón látható szobában. A kutyák csak akkor tudták konzisztensen megtalálni a tárgyat amikor a rejtekhelyek ugyan abban a szobában voltak mint a videó demonstráció, de nem ha a rejtekhelyek egy másik szobában voltak. A vizsgálat eredményeiből arra következtettünk, hogy a kutyák nem tettek különbséget a valóság és a videón láttottak között és nem értették meg a videó képek referenciajának jellegét.

A harmadik vizsgálatban egy a második vizsgálatéban alkalmaztattott alkalmazatban hasonló tárgykeresési teszben figyeltük meg, hogy a kutyák különbözőképpen reagálnak-e egy fejjel lefele vagy fejjel felfele megjelenített emberre. A kutyák projektor vászon iránti orientációjának és helyes választásainak elemzése arra utalt, hogy a fejjel lefele kondícióban kezdetben a videót csak mint egymástól független vizuális ingerek összességét érzékeltek és csak a későbbi próbák során ismerték, fel hogy az egy embert ábrázol.

A negyedik vizsgálatban virtuális tárgykeresési tesztek végeztünk a kutyákkal az érintő képernyős módszer segítségével. Összehasonlitva a virtuális teszt különféle változataiban kapott eredményeket korábbi, való életben végzett tesztek eredményeivel arra a következtetésre jutottunk, hogy a kutyák feltehetőleg nem ugyan azon kognitív képességek segítségével oldották meg a virtuális, mint a valós feladatokat.
I. A doktori értekezés adatai
A szerző neve: Péter András
MTMT-azonosító: 10028302
A doktori értekezés címe és alcíme: Evaluating the potential of artificial visual stimulus presentation in object search tasks for dogs
DOI-azonosító: 10.15476/ELTE.2016.145
A doktori iskola neve: Biológia Doktori Iskola
A doktori iskolán belüli doktori program neve: Etológia Program
A témavezető neve és tudományos fokozata: Dr. Pongrácz Péter, egyetemi docens
A témavezető munkahelye: ELTE TTK Etológia Tanszék,
1117 Budapest, Pázmány Péter sétány 1/C

II. Nyilatkozatok
A doktori értekezés szerzőjeként
a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a Természettudományi Kar Dékáni Hivatalának Doktori, Habilitációs és Nemzetközi Ügyek Csoportja ügyintézőjét, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltsön a feltöltéshez szükséges nyilatkozatokat.
b) kérem, hogy a mellékelt kérelemen részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;
c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (dátum)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;
d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéségig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adattait tegyen közzé. Ha a könyv a fokozatszerzést követően egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésen és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.

2. A doktori értekezés szerzőjeként kijelentem, hogy
a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti , önálló szellemi munkám és legjobb tudomáson szerint nem sértem vele senki szerzői jogait;
b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.
3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: Budapest, 2016 09 29

[Untemplated signature]

…………………………………………
a doktori értekezés szerzőjének aláírása