From the Messerli Research Institute
University of Veterinary Medicine, Vienna
Unit of Comparative Cognition
(Head: Univ.-Prof. Mag. Dr.rer.nat. Ludwig Huber)

Food vs. Toy - Exploring How Reward Type Shapes Dog Search Behaviour

Master Thesis
Submitted in partial fulfilment of the requirements for the degree of
Master of Science (M.Sc.)
at the University of Veterinary Medicine, Vienna

by
Christiana Tsiourti

Vienna, September 2023

Supervisors: Priv.-Doz. Dr. Zsófia Virányi

Reviewer: Ass. Prof. Dr. Friederike Range

Acknowledgments

I want to extend my heartfelt gratitude to the incredible people who supported me throughout this master's program. The friends, family, and colleagues mentioned here (and many more!) played a vital role in making this possible.

First and foremost, I am truly thankful to my life partner, Sten, who encouraged me to pursue my passion for animal research and lovingly stood by me throughout the completion of this degree! Without your unwavering support, none of this would have been possible.

I would like to express my gratitude to my exceptional supervisor, Zsófia Virányi. Right from the beginning, she welcomed me warmly to Messerli and provided constant support, motivation, and inspiration. She has been the best supervisor one could ask for, and I eagerly anticipate our future collaborations!

I owe a big thank you to my colleague and dear friend, Katie Baynham. Working alongside her on these experiments has been an absolute pleasure. Her meticulous attention to detail, hard work, and dedication were pivotal to the success of this project.

I am deeply grateful to Livia Langner and Miriam Ross for their enthusiastic involvement in this project and the countless hours they dedicated to recruitment, conducting tests, and coding data.

Karin Bayer, your dedication to the Clever Dog Lab and your willingness to always lend a helping hand are immensely appreciated.

Anna Zamansky and Nareed Hashem, thank you for the excellent collaboration on the automatic behavioural analysis part of the project.

Remco Folkertsma and Roger Mundry, thank you for your invaluable assistance with the challenging statistical aspects of the project.

Bettina, thank you for the peaceful and inspiring days I spent in your beautiful office while writing.

To my fellow IMHAIs, thank you for the fun and supportive companionship throughout the entire program. It has been a pleasure getting to know each one of you!

To my loving family - mom, dad, and Vaggelis, thank you for always being there for me. Additionally, to my extended family, Viola, and Gunther, your support has been invaluable to the completion of this work.

Last but not least, I extend my appreciation to all the dogs and their owners who generously participated in our experiments. Your participation was essential to our work, and I am grateful for your contribution.

To all those mentioned above, your support, dedication, and collaboration have been instrumental in the completion of this master's, and I am deeply grateful for your help and encouragement. Thank you!

Plagiarism disclaimer

I hereby declare that this thesis is my own and autonomous work. All sources and aids used have been indicated as such. All texts either quoted directly or paraphrased have been indicated by in-text citations. Full bibliographic details are given in the reference list, which also contains internet sources containing URL and access date. This work has not been submitted to any other examination authority.

Graz, 10.08.2023

Christiana Tsiourti

Christiana Tsiourti

Abstract (English)

Domestic dogs are renowned for their social cognition and proficient interactions with humans, yet comparatively little is known about how they represent and reason about the physical world around them. Specifically, while it is well-established that dogs do have an object permanence (*i.e.,* can form mental representations of objects in their environment), it remains unclear how they represent different types of rewards, and how the nature of those rewards impacts their behaviour and performance in experimental tasks. This thesis explores two competing hypotheses, namely (a) that different types of rewards trigger the use of distinct cognitive strategies (inference and association) during object search, and (b) that dogs hold distinct mental representations for different types of rewards. Two groups of dogs are tested in a meticulously designed variation of the classic Piagetian invisible displacement task – one group with food and another with a toy as reward. Dogs see the reward (food or toy) “disappear” from a displacement device, after passing behind three hiding screens. Dogs are allowed to search for the reward, and their speed as they move between screens is measured. If dogs use inference by exclusion to find the reward, they should speed up after finding two empty screens, indicating their confidence in finding the reward behind the third screen. In contrast, the experiments yielded robust evidence that dogs tended to slow down as they approached the final screen, signifying a predominant reliance on associative strategies rather than inferential reasoning. Importantly, the reward type, be it food or a toy, did not significantly affect their search behaviour. That is, in this study we found no evidence that dogs would represent food and toy rewards differently or that these rewards would activate different cognitive strategies in dogs. These findings hold significance from practical and theoretical standpoints, and lay a solid methodological groundwork for further research distinguishing between behaviour rooted in logical inference and associative generalization across various species.

Abstract (German)

Hunde sind bekannt für ihre soziale Kognition und ihre geschickten Interaktionen mit Menschen. Dennoch ist vergleichsweise wenig darüber bekannt, wie sie die physische Welt um sich herum wahrnehmen und in ihr interagieren. Insbesondere ist es zwar bekannt, dass Hunde über Objektpermanenz verfügen (d.h., mentale Vorstellungen von Objekten in ihrer Umgebung bilden können), aber es bleibt unklar, wie sie z.B. verschiedene Arten von Belohnungen wahrnehmen und wie die Art der Belohnungen ihr Verhalten und ihre Leistung in experimentellen Aufgaben beeinflusst. Diese Dissertation untersucht zwei konkurrierende Hypothesen, nämlich (a), dass verschiedene Arten von Belohnungen die Verwendung unterschiedlicher kognitiver Strategien (Inferenz und Assoziation) während der Objektsuche auslösen, und (b), dass Hunde unterschiedliche mentale Vorstellungen für verschiedene Arten von Belohnungen haben. Zwei Gruppen von Hunden werden in einer Variation der klassischen Piaget'schen Aufgabe zur nicht sichtbaren Verschiebung getestet - eine Gruppe mit Nahrung und eine andere mit einem Spielzeug als Belohnung. Die Hunde beobachten, wie die Belohnung (Nahrung oder Spielzeug) hinter drei Versteckschirmen verschwindet. Die Hunde dürfen nach der Belohnung suchen, und ihre Geschwindigkeit, während sie sich zwischen den Bildschirmen bewegen, wird gemessen. Wenn Hunde Inferenz durch Ausschluss verwenden, beschleunigen sie nach dem sie bereits zwei leere Bildschrime vorgefunden haben, da sie nun sicher scheinen die Belohung hinter dem dritten Bildschirm vorzufinden. Im Gegensatz dazu lieferten die Experimente robuste Beweise dafür, dass die Hunde dazu neigten, sich zu verlangsamen, wenn sie sich dem letzten Bildschirm näherten, was auf eine überwiegende Abhängigkeit von assoziativen Strategien hinweist, anstatt von inferentialem Denken. Wichtig ist, dass die Art der Belohnung, sei es Nahrung oder ein Spielzeug, ihr Suchverhalten nicht signifikant beeinflusste. In dieser Studie fanden wir also keine Hinweise darauf, dass Hunde Nahrungs- und Spielzeugbelohnungen unterschiedlich repräsentieren würden oder dass diese Belohnungen verschiedene kognitive Strategien bei Hunden aktivieren. Diese Erkenntnisse sind von praktischer und theoretischer Bedeutung und legen eine solide methodische Grundlage für weitere Forschungen zur Unterscheidung zwischen Verhalten, das auf logischer Inferenz und assoziativer Verallgemeinerung bei verschiedenen Arten basiert.

Table of Contents

1. General Introduction

In studies of animal cognition, researchers employ diverse rewards to motivate and reinforce animal behaviour. Some of the most common types of rewards include food, social interactions (*e.g.,* praise, petting, or play with a human handler or conspecific), object rewards (*e.g.,* toys or novel items), sensory rewards (*e.g.,* auditory, or olfactory stimuli), and access to preferred environments (*e.g.,* outdoor enclosure or a specific location). The choice of reward type depends on the species being studied, the task requirements, and the preferences of the animals involved. The perceived value of a reward significantly impacts the level of interest and motivation with which animals engage in a task (*e.g.,* (Belger & Bräuer, 2018; Krichbaum & Lazarowski, 2022)). Furthermore, different rewards can have distinct effects on high-level cognitive processes ranging for inhibitory control (*i.e.,* the ability to suppress an impulsive response in favour of one that is more productive), to learning, mental representation, and reasoning (*e.g.,* (Kruijt, 1964; Sulikowski & Burke, 2011)).

1.1. Reward and Incentive Motivation

The majority of dog studies employ either food or toys as rewards, often selecting the preferred reward based on owner reports. Because of the heightened arousal caused by food, a prevailing assumption is that animals perform better in problem-solving tasks when non-food items are used as rewards. However, remarkably few studies have systematically examined the impact of food versus toys on dogs' problem-solving abilities. Existing findings in this domain offer a mixed perspective. (Krichbaum & Lazarowski, 2022) compared the performance of dogs in the cylinder task when a treat or a ball is used as a reward. The type of reward significantly affected dogs' ability to correctly detour, and dogs **performed better when solving the task for a treat rather than a ball**. The authors of this study, in contrast with the above assumption, argued that the high desirability of the ball likely induced heightened arousal and diminished inhibitory control. In a different study where dogs had to locate a hidden reward behind one of two fences, **dogs more frequently chose the correct fence when rewarded with a toy**. When the baiting process was not visible, toy-rewarded dogs also sought additional information (by checking through a gap in the fence) more frequently than food-rewarded dogs (Belger & Bräuer, 2018). When the reward was food, dogs may have been more impulsive and therefore less likely to make correct choices or show metacognitive abilities (but see Section 1.2 for a discussion on alternative explanations). Other studies have also shown that dogs behave more impulsively when a more highly preferred reward is used (Brucks et al., 2017).

Reward value can be altered either by changes in quantity or by using a reward of different quality. **Dogs can discriminate gross quantitative differences in food rewards** and **show a preference for more numerous rewards** (Ward & Smuts, 2007). However, little is known about how differences in reward quantity impact their behaviours related to response time and effort (Bensky et al., 2013). (Leonardi et al., 2012) measured dogs' willingness to wait to exchange a reward with an experimenter and manipulated the quantity of the exchange reward. The **larger the quantity of the exchange reward, the longer dogs were willing to wait**.

A study by (Riemer et al., 2018) demonstrated that higher quality food rewards entailed greater incentive motivation in dogs than a greater quantity of a lower value reward. Specifically, dogs ran significantly faster for higher quality food compared to lower quality food (1 piece of sausage vs 1 piece of dry food). In contrast, there was no significant effect of food quantity (1 vs 5 pieces of dry food) on their running speed. (Bräuer & Call, 2011) found that when tested in a violation of expectation paradigm, **dogs searched longer and remained in proximity to the experimenter if they recovered food that was of different type or quality from what they previously observed being hidden in a container** (*i.e.,* high value sausage vs. low value dry food). This behaviour suggests that dogs did not expect to find just any reward, but for the specific reward that they had seen being hidden. However, when (Bräuer & Call, 2011) tested dogs in the aforementioned two-screen task with food rewards of varying quality (*i.e.,* high-value sausage vs. low-value dry food), **dogs' performance was unaffected, though they were notably quicker in retrieving the high-quality food reward compared to the low-quality one**. (Brucks et al., 2017) tested dogs in a delay gratification task (*i.e.,* a task which requires individuals to choose between an immediate option of lower value and a delayed option of higher value). Food was used as reward and was systematically varied across a quality and a quantity condition. Dogs tolerated higher delays for the higher quantity of the less preferred reward than for the higher quantity of the more preferred reward type. This indicates that **using the highly preferred food as the immediate reward caused dogs to act more impulsively, whereas using it as the delayed option facilitated waiting.**

The aforementioned findings underscore the distinct impact of rewards on animal motivation and inhibitory control (*i.e.,* the ability to restrain impulsive responses in favour of more productive actions). In light of these findings, researchers have started exploring whether the use of tokens (inherently non-valuable objects exchangeable for valuable rewards) can alleviate issues arising from the stimulating nature of rewards (particularly food) and enhance animal task performance. Chimpanzees tested in a reverse–reward contingency task (*i.e.,* an inhibition task where subjects presented with two different quantities of items must select the smaller quantity to obtain the larger reward) could not inhibit their motivation to choose the larger quantity when tested with food stimuli. However, when tested with Arabic numerals, they succeeded by selecting the smaller numeral, which led to the larger reward. By contrast, symbolic stimuli closely resembling the food, such as rocks having a one-to-one correspondence with food, did not ameliorate performance (Boysen, 2012). (Fernand et al., 2018) found that dogs tested in a reverse–reward contingency task chose optimally when analogous symbolic stimuli were used (*i.e.,* they selected the larger stimulus to receive the large-magnitude reward), but, in contrast to chimpanzees, they failed to reverse their choices to receive optimal rewards when the consequences of their choices were reversed (*i.e.,* the smaller stimulus led to a large-magnitude reward, and the larger stimulus to a small-magnitude reward). It remains dubious whether the use of tokens, or symbolic stimuli allows dogs to inhibit strong behavioural predispositions towards food, thus improving their performance. Surprising, (Range et al., 2011) found that **dogs performed significantly better in a means-end task if food was used instead of a toy object that could be exchanged for a food reward** (*i.e.,* a token). The authors proposed that the heightened motivation and overall excitement observed in dogs rewarded with toys may have superseded their problem-solving capabilities. This hypothesis finds support in several aspects of the study: (a) the subject pool exclusively included dogs with a strong inclination to retrieve toys, suggesting a predisposition towards heightened motivation in this subgroup, (b) in between task trials, the dog owners engaged in brief play sessions with dogs using the retrieved toy, to sustain the dogs' motivation levels throughout the experiment, and (c) the dogs rewarded with toys required fewer sessions to successfully complete the task when compared to dogs receiving food rewards.

1.2. Reward and Cognitive Strategies

Going beyond motivation and inhibitory control, other studies suggest that **when working for diverse reward types, animals rely on varying cognitive strategies, which can significantly influence their behaviour and performance**. An early example of reward-dependent learning is the work of (Kruijt, 1964), who attempted to train male Burmese Red Junglefowl to open a door using courting behaviours. Interestingly, conditioning was successful only when a female hen was situated behind the door, whereas food behind the door food yielded no effective learning. In another illustrative example, (Sulikowski & Burke, 2011)showed that noisy miner birds engage different cognitive strategies in response to being rewarded with invertebrates or nectar in laboratory tasks. When rewarded with nectar (an immobile, readily depleted resource), birds used a spatial working memory strategy that involved spontaneous encoding of point location information. In contrast, when rewarded with invertebrates (a mobile resource, not spread out across the environment), birds used a spatial memory-based strategy with attention to point location information. Collectively, these studies, underscore how the type of reward can intricately modulate an animal's learning, decision-making processes, and overall cognitive performance.

With respect to dogs, the aforementioned study of (Belger & Bräuer, 2018) highlights the differential effects of food versus toy rewards on the metacognitive skills of dogs during a search task. In three consecutive experiments, dogs had to find a reward that was hidden behind one of two V-shaped screens. Each screen had a small gap, which allowed dogs to seek additional information about the area behind the screen. Various parameters were manipulated to investigate their effects on the dogs' searching behaviour: the type (food vs. favourite toy) and quality (high vs. low) of the reward, the time delay between baiting and choosing, and whether dogs had visual access to the baiting procedure or not. Interestingly, dogs rewarded with toys selected the correct fence more often and tended to seek additional information by checking the gap on the screen more often than dogs rewarded with food.

The findings of (Belger & Bräuer, 2018) provide evidence that dogs search in a more flexible way when they are rewarded with a toy. As discussed in the previous section, the greater performance with the toy reward might be explained by daily life experience with toy-search, or might be correlated with motivation (*i.e.,* when rewarded with a toy they were not as over-motivated as with food, thus better able to inhibit a direct approach before checking for additional information**). Nevertheless, it is also possible that the nature of the reward affected other, higher level cognitive processes like reasoning or mental representations.** Specifically, the authors postulate that **dogs may use different cognitive strategies with different rewards**, and toy-search may rely on an inferential mechanism more than food-search. Dogs also showed flexible searching strategies (*i.e.,* inferential reasoning) when they searched for a toy in previous studies (see Section 1.4). Another alternative hypothesis they propose is that **dogs may mentally represent the two rewards and their locations differently**. A favourite toy is a concrete object – **so toy-search might be more focused** since dogs know exactly what they are searching for. In contrast, there could potentially be multiple pieces of food around, so **food-search might be more diffuse**. The methodology of (Belger & Bräuer, 2018) could not disentanlge between these hypothesis – so it remains unclear why dogs performed better with the toy reward.

1.3. Dogs’ Mental Representation of Hidden Objects

Search tasks in which an animal is requested to find a reward that is hidden out of sight, rely on the concept of ***object permanence*** – *i.e.,* the ability of an animal to remember and mentally represent objects outside of their field of perception (Piaget, 1952). Developmental psychologist Jean Piaget described the development of object permanence in human infants as having six stages (Piaget, 1952). **Stage 4** represents the most elementary understanding of object permanence. At this stage, subjects are capable of retrieving a target object that has been fully hidden from their view when they have directly witnessed the hiding. At **Stage 5a**, subjects perform well on multiple trials of sequential **visible displacement** in which the hiding location is different between trials. In **Stage 5b**, subjects are successful at **successive visible displacement** trials where the target object is not only visibly displaced behind one screen per trial, but an experimenter (visibly) moves the object behind multiple screens before leaving it behind the last visited screen. To reach **Stage 6**, subjects have to successfully solve **invisible displacement tasks** in which the hiding of a target object is not directly witnessed. Instead, subjects see the target object being placed inside a displacement device (*e.g.,* opaque container) which is then moved behind or under multiple alternative hiding places. Once all displacements are concluded, the displacement device is shown to be empty. To solve the task, the subject must reason by exclusion from the evidence of the empty displacement device and infer the target object’s current location based on its past series of movements.

Canine research on object permanence indicates that dogs are capable of forming mental representations of objects in their environment and understand that objects still exist even when they have left their immediate perceptual field (for an overview see (Range & Marshall-Pescini, 2022b), and (Bensky et al., 2013)). There is consistent evidence that dogs are successful in Stage 5 visible displacement tasks, but it remains controversial whether they can reach Stage 6 of Piaget’s model. Early studies by (Triana & Pasnak, 1981) found that dogs were able to successfully solve the **successive visible displacement task**, even when controlling for olfactory cues. These findings have been repeatedly replicated (Fiset & Plourde, 2013; Gagnon & Doré, 1992). (Triana & Pasnak, 1981) initially claimed that dogs could successfully solve successive invisible displacement tasks. However, these findings have not been robustly replicated. In fact, more recent studies revealed that dogs who seem to pass **invisible displacement tasks** may do so by relying **on associative learning without having to mentally represent the unperceived displacement of the target object** – shown by the fact that they fail the task when strict controls are introduced in the methodology (e.g. (Collier-Baker et al., 2004; Watson et al., 2001)). There are at least three associative strategies that, if not controlled, could account for success on invisible displacement tasks without relying on inferential reasoning: local associative rules (*e.g.,* position of displacement device), sensory cues, and social cues (for details see (Jaakkola, 2014). Through a series of experiments with strict controls for associative cues, (Collier-Baker et al., 2004) gathered compelling evidence indicating that the search behaviour of dogs in Stage 6 invisible displacement tasks is primarily guided by simple associative rules rather than a mental representation of the object’s past trajectory.

1.4. Inferential Reasoning in Dogs’ Search for Objects

As discussed above, the dogs’ performance in the Piagetian invisible displacement task appears to best be explained by alternative strategies based on associative learning and environmental cues, instead of inferential reasoning (Collier-Baker et al., 2004; Watson et al., 2001). **In tasks that involve a choice between several alternatives, inferential reasoning by exclusion is a cognitive process in which the correct choice is deduced by logically excluding other potential alternatives (Call, 2006).**

For example, if one knows that a reward is hidden at one of two locations, A or B, then one can logically infer from its absence at location A that it must be at location B. **(Watson et al., 2001) used a variation of the invisible displacement task to test the ability of inferential reasoning by exclusion in dogs and 4- 6-year-old children**. Subjects saw a toy "disappear" from a displacement device, after passing behind three screens. Subjects were allowed to search for the toy, and their speed as they moved from screen to screen was measured. The authors reasoned that if subjects used inference by exclusion (the toy is in location 1, 2 or 3), once they discovered two empty screens, they would speed up, reflecting increased confidence that the toy must be in the third. The findings revealed significant differences in how children and dogs approached the task. While the **children reliably increased the speed of their approach toward the final screen, the dogs became slower**, as if an associative mapping between the reward and the screens was being extinguished. This behaviour was construed as evidence that **dogs' search was not driven by the same logical reasoning as in children, but rather, they searched based on screen-reward associations formed due to their immediate sensory experiences of the container approaching the screens**.

In contrast to the above-mentioned findings, evidence from another experimental paradigm - the cup task (Call, 2006), reveals that, **in certain conditions, dogs have the ability for inferential reasoning by exclusion – particularly when searching for hidden toys** (Erdohegyi et al., 2007). In the cup task, subjects are presented with two cups, one of which has been surreptitiously baited with a reward. When the interior of the empty is shown, or the cup is shaken, subjects are expected to infer by exclusion that the alternative cup contains the reward. While dogs generally preferred the cups that the experimenter manipulated (Bräuer et al., 2006; Erdohegyi et al., 2007), they solved the task using inference by exclusion after controlling for local enhancement. When the experimenter touched and looked at both cups but revealed only the content of one of them (baited or non-baited), dogs picked the baited cup (Erdohegyi et al., 2007). When tested in an auditory-only version of the task (Bräuer et al., 2006), dogs performed poorly: when the empty cup was shaken, they selected the baited cup on less than 40% of trials.

Positive results from the cup task must be interpreted with caution because successful performance can be achieved using low- and high-level cognitive strategies (Völter & Call, 2017). Subjects may solve the task simply by learning to avoid the empty cup, without any knowledge about the other rewarded cup (low-level strategy). Other subjects, however, may be capable of inferring the presence of the reward under the non-lifted cup (high-level strategy). The invisible displacement task is a better alternative, since subjects do not see any of the hiding places being empty, thus the explanation of avoiding the empty location is less likely.

Another paradigm for testing inference by exclusion is matching-to-sample. In this task, subjects are trained on conditional discrimination, and then a novel sample is introduced, with a choice between a novel and a familiar comparison. **(Aust et al., 2008) found evidence of reasoning by exclusion in pet dogs using a touchscreen version of this task**. There is also some sporadic evidence (based on the performance of single individuals) of inferential reasoning from fetching paradigms where Border Collies acquired the relation between a word and a referent word, and could later infer the referent of new words by exclusion learning ((Kaminski et al., 2004; Pilley & Reid, 2011)).

In sum, evidence of the inferential abilities of dogs in physical problem-solving tasks is scarce. Based on the poor performance of dogs in invisible displacement tasks, and other tasks in the physical domain (e.g. means-end understanding tasks), it has been proposed that the domestication process might have had a detrimental effect on their physical cognition (Frank, 1980) while increasing their social cognitive skills (Miklósi et al., 2003). Although evidence for the presence of inferential reasoning capacity in dogs is scarce, this does not necessarily imply the absence of this ability altogether. Associative learning is often placed in opposition to forms of reasoning, to determine which cognitive process enables animals to solve problems. Nevertheless, there is also **the possibility that associative and inferential processes are complementary, and used flexibly, by different individuals, in different situations to facilitate knowledge acquisition and problem-solving.** Hence, dogs' inability to perform well on invisible displacement tasks when associative cues are controlled might stem from **natural variation** (*e.g.,* some dogs are more prone than others to using particular cognitive strategies), or **individual strategy preferences** (*e.g.,* some dogs choose to use associative learning while others use inferential reasoning). Another possibility, examined in this thesis, is that inferential reasoning and associative learning are complementary strategies, with prevalence of one or the other in a task being driven by the **nature of the reward**. This hypothesis was proposed by (Belger & Bräuer, 2018), but their the methodology did not allow for testing it.

1.5. Aims and Hypotheses

While it has been reliably established that dogs are capable of creating mental representations of objects in their environment and understand that objects still exist even when they have left their immediate perceptual field, it remains unclear how they represent different types of rewards, and whether, or under which conditions, they make use of inferential reasoning when they search for them. This thesis aims to further investigate the effects of reward type on the cognitive performance and search behaviour of dogs. Specifically, the main research question addressed here is the following: "***Do dogs search for hidden food and toy rewards differently, and, if so, why?***".

Building upon the insights provided by (Belger & Bräuer, 2018), two alternative hypotheses are proposed to answer this question:

- **Hypothesis (H1):** Dogs may rely on ***different cognitive strategies*** when searching for different types of rewards, and toy-search may rely more heavily on inferential reasoning than food-search does[1].

- **Hypothesis (H2):** Dogs may hold ***different mental representations*** for different types of rewards. A toy is a concrete object, and dogs likely know that it can be found only in one location. In contrast, food is a resource that is typically abundant in the environment, and there is possibly more than one piece around. Therefore, food, and its location, may be represented in a more diffuse manner than a toy.

To differentiate between the use of different cognitive strategies (H1) and different mental representations (H2), the methodology put forth by Watson et al. (2001), herein referred to as the *Watson task*, is used. The *Watson task* is a variation of Piaget's *invisible displacement* paradigm (Piaget, 1952) which involves showing participants an object that is subsequently hidden and moved to a new location. Participants are then presented with a choice of screens or containers to find the hidden object. The challenge lies in the fact that the object is moved to a different, unseen location, requiring participants to infer its new position.

---

[1] This hypothesis assumes that a) both inferential reasoning and association may be available to dogs, and b) the use of inferential reasoning is not a binary trait but rather occurs along a continuum.

In the *Watson task*, an experimenter displayed subjects a reward (in the original study always a toy) which was subsequently concealed within a container and moved behind three different screens. The experimenter crouched behind each screen in succession, then returned, and presented an empty container. Subjects were allowed to search for the reward, and they were timed as they approached the first, second, and third screen. The novelty, and key element of the task, was the fact that, while crouching behind one of the screens, the experimenter discreetly concealed the object within their pocket, thus making it impossible to find it behind any of the three screens.

With this procedure, subjects were faced with a scenario where *´the object is behind neither screen 1 nor screen 2.´* The authors reasoned that if subjects relied on inferential reasoning, more specifically reasoning by exclusion (*i.e., the object is behind screen 1, 2, or 3*), they should *deduce* that the object must be behind screen 3, as it is the only option left after excluding screens 1 and 2. As a result, they should *accelerate* upon nearing the final screen after their unsuccessful attempts to find the object behind the previous two screens.

The *Watson task* stands out as an insightful tool that can distinguish between behaviour that reflects logical inference from behaviour that reflects associative generalization. Hence, the primary aim of this thesis is to replicate the task with two different types of rewards (food and toy), to investigate whether and how they impact the search behaviour of dogs. Beyond exploring the effects of reward, a secondary aim of this thesis is to ascertain the replicability of Watson et al.'s (2001) findings, with a more rigorous behavioural measurement protocol.

Upon a meticulous examination of the methodology of Watson et al. (2001), a notable limitation to their original conclusion emerged. The core behavioural measurement employed in the analysis was the subjects' *speed* while approaching the first, second, and third screen. *Speed* was defined as a composite measure, encompassing both travel time between screens, and time spent exploring each screen. This composite speed measure raises concerns as it may inadvertently conflate the *slowing down* behaviour observed in dogs. In other words, dogs might seemingly slow down when approaching the final screen not due to slower travel speed but rather because they invested more time exploring the later screens. Such a scenario may lead to erroneous conclusions about the cognitive strategy of dogs. Extended exploration behind later screens may, in fact, signify a growing anticipation of locating the reward, serving as a clear indicator of the use of reasoning by exclusion (*i.e., if the object is not in screen 1 and not in screen 2, then it must be in screen 3*).

To rectify this noteworthy limitation, the present thesis endeavours to refine the methodology of Watson et al.'s (2001), with a distinct emphasis on analyzing travel time between screens (defined as *pace*), and time spent exploring each screen (defined as *exploration time*), as two discrete behavioural measurements. This critical differentiation promises a more comprehensive analysis of dogs' search behaviour, thereby facilitating more precise conclusions regarding the utilization of inference versus association in their decision-making processes.

1.6. Measurements and Predictions

To address our two research questions, four predictions are defined, centred on the two following behavioural measurements:

(1) **Pace**: In the context of this thesis, *pace* is defined as the inverse of *speed* as defined by Watson et al. (2001). It refers to the travel time divided by the distance covered by dogs, when approaching their first, second, and third screen choices (referred to as P1, P2, and P3 respectively). Whereas speed is measured in units of distance per unit of time, pace is measured in units of time per unit of distance. The pace measurement allows for a comparison between dogs regardless of the distance they covered or the routes they took during each travel. This adjustment accommodates diverse search behaviours and ensures that the pace metric remains meaningful even if dogs use different paths to approach the screens.

(2) Exploration time: This refers to the time spent by dogs exploring the first, second, and third screen they approached (referred to as E1, E2, and E3 respectively).

We used *pace* to (1) investigate whether Watson et al. (2001) were right and whether dogs indeed slow down during the search of the three screens, and, if this is the case, to (2) test our H1 and H2 that predict opposite effects between the food and toy groups, as outlined below.

- **Prediction 1:** As per H1, we anticipate a **stronger slowing down effect in the food group compared to the toy group**. This is because the food search relies less on inferential mechanisms compared to the toy search. See Figure 1(a) for a visual representation.

- **Prediction 2:** As per H2, we anticipate a **weaker slowing down effect in the food group compared to the toy group**. This is because the diffuse nature of food in the environment maintains a strong association between the reward and hiding location, even after encountering empty screens. In other words, dogs retain their expectation of finding the food reward throughout the search, and this is manifested as a less pronounced decrease in pace compared to the toy search. See Figure 1(b) for a visual representation.

In terms of *exploration time*, H1 and H2 predict opposite behaviours.

- **Prediction 3:** As per H1, we anticipate a **progressive increase in exploration time in the toy group compared to the food group**. This is because the toy group will display a rising sense of surprise as they fail to locate the object behind each screen, owing to their heightened dependence on inferential mechanisms. See Figure 1(c) for a visual representation.

- **Prediction 4:** As per H2, we anticipate **overall longer exploration times for all the screens in the food group compared to the toy group**. This is because the diffuse nature of food in the environment maintains a strong association between the reward and hiding location, even after encountering empty screens. In other words, dogs retain their expectation of finding the food reward throughout the search, and this is manifested as prolonged exploration. See Figure 1(d) for a visual representation.

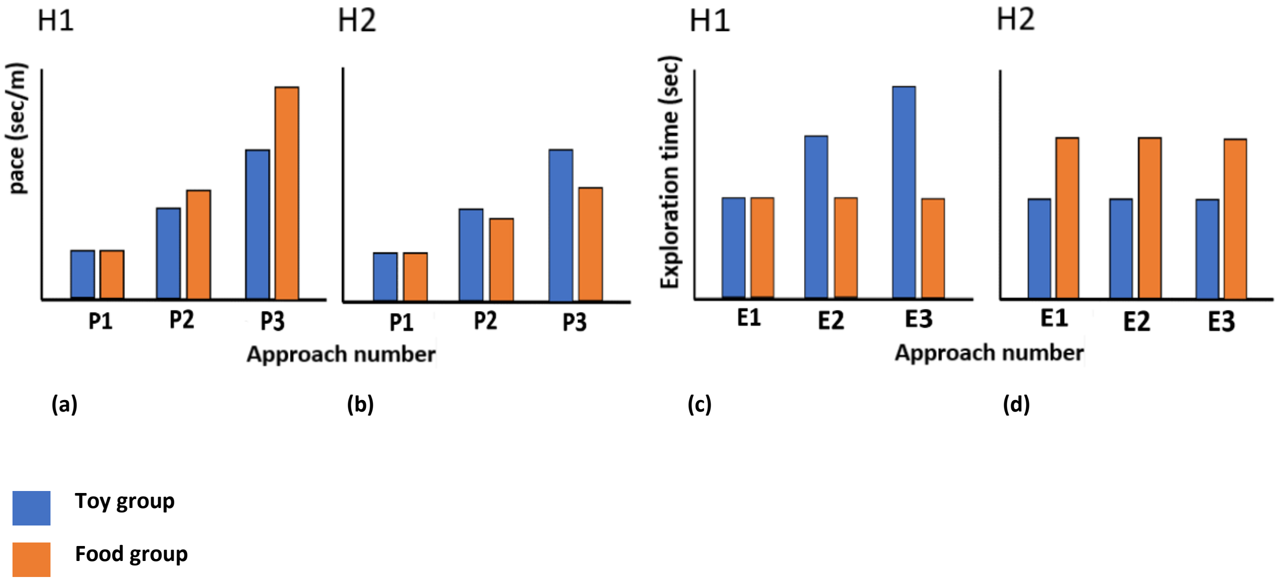

Figure 1: Graphical representation of predictions 1-4 (a-d) for average pace (*left*) on approaching first, second, and third screen choice (P1-P3, respectively) and average exploration time (*right*) of first, second, and third screen choice (E1-E3, respectively), for toy group (blue) and food group (orange).

2. Experiment 1 – The Watson Task

2.1. Ethical Statement

The experiments herein reported (Experiment 1, Experiment 2) were approved by the Vienna University of Veterinary Medicine ethics and animal welfare committee following GSP guidelines and Austrian legislation (approval number: ETK-159/10/2020). Written consent to participate in the study was obtained from the dogs' owners. In addition, dog owners, and staff involved in the study signed a data protection consent form concerning the collection of video material, pictures, and personal details.

2.2. Subjects

The study was advertised through digital platforms and word of mouth. To participate, dogs had to be (a) pets, (b) at least 1-year-old, and (c) motivated to search for either food or toys. Food motivation was defined as approaching and eating a piece of food consecutively, and toy motivation was defined as approaching and/or retrieving a toy consecutively (for details see Section 2.4). In total, 56 dogs were recruited (27 males, 29 females, age: mean = 5,48, SD =3,34) naïve to the experimental procedure (for demographic details see Appendix A). We asked the dog owners whether their dog was more motivated to search for food or a toy and assigned dogs to one of two corresponding groups (between-subject design). The food group included 29 dogs (11 males, 18 females, age: mean = 6,62, SD =3,33), and the toy group included 27 dogs (16 males, 11 females, age: mean = 4,26, SD =2,90). We counterbalanced the groups as much as possible concerning age, sex, and breed.

2.3. Experimental Setup

Testing took place at the Clever Dog Lab (Vienna University of Veterinary Medicine) in a bare well-lit room (size: 7,25m x 6m). The experimental setup is shown in Figure 2. Three yellow wooden screens (sides: 70 cm wide X 90 cm high) were positioned along a semicircle at an equal distance from a starting point, where the owner was seated holding the dog in front of them. The screen arrangement ensured that (a) dogs were placed at an equal distance from all the screens, and (b) while searching, dogs' line of sight was restricted at most to one screen's hiding area at a time.

A plastic, opaque flowerpot (16cm deep, 18cm diameter) was used as a displacement device. For the toy group, the reward was the dog's favourite toy (*e.g.,* ball), provided by the owner (see Figure 3 for pictures of the rewards). Two criteria were specified for the toy: (a) the toy had to be small enough to be fully hidden inside the displacement device, and (b) the toy should not make any noise while being handled. For the food group, the reward was a plastic, green opaque bowl containing one piece of the dog's preferred food (*e.g.,* sausage). The bottom of the bowl and the flowerpot were lined with polystyrene, to absorb any sound made as the objects were deposited behind the hiding screens. Tape on the floor marked the positions of the wooden screens, the subject's starting position as well as the positions of the experimenter and owner. Water was available ad libitum.

The experimental set-up followed that of (Watson et al., 2001), but there were some differences. Namely, in this study, the testing room was larger, and the screens were taller than the ones used in the original study. By using taller screens, we made sure that the experimenter was fully occluded during the hiding process – and thus eliminated the transmission of any inadvertent cues that might guide the search behaviour of the subjects.

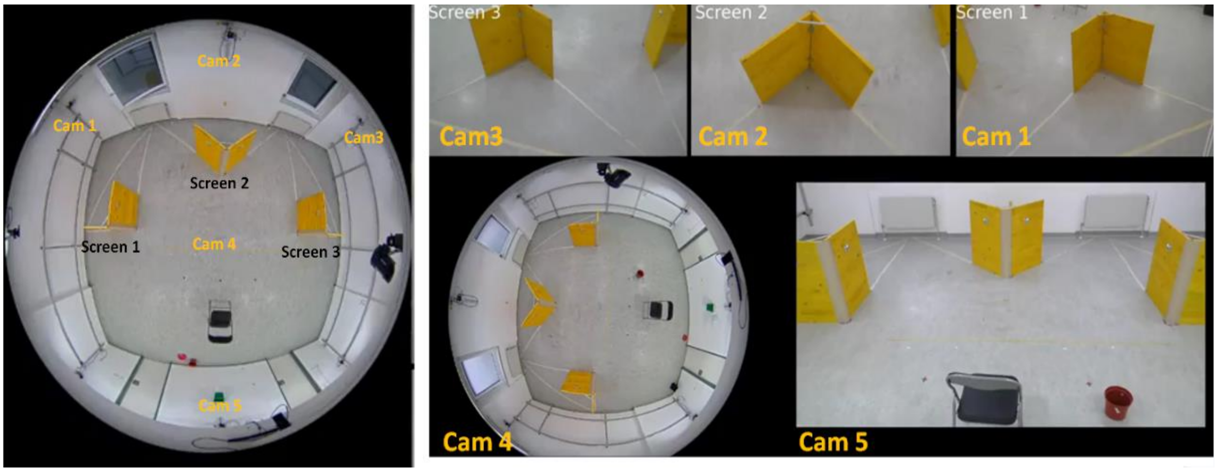

Figure 2: Photograph of the experimental setup in the testing room showing the position of the three wooden screens along a semi-circle (left), and the view of the setup from the five cameras used to record the experiment (right).

2.4. Experimental Procedure

The test was administered within a single session lasting approximately 30 min on average. Dogs were tested individually with their owner present. Three different experimenters ran the tests. The experimental procedure is outlined below.

Screening Trials: A pre-condition for this test was that dogs show high and sustained motivation for the reward. Thus, an initial screening procedure, lasting about 5 min, was used to confirm that dogs were motivated to retrieve the toy (toy group) or to eat the food (food group). The screening took place in another, similar room of the Clever Dog Lab. The owner sat on a chair with the dog in front of them. If needed, owners manually held their dogs in position by the collar or shoulders to prevent them from initiating any movement ahead of time. For the food group, an empty plastic, opaque bowl (which later served as the reward) was placed in front of the dog, at two meters. The experimenter walked up to the bowl, dropped in a piece of the dog's favourite food, and returned beside the owner. In the first three trials, the dog could see the experimenter baiting the bowl (*visible baiting*); in the next seven trials, the experimenter picked up the bowl with their back turned to the dog, and "*secretly*" dropped the food inside (*invisible baiting*). A salient phrase such as "*go*" or "*search*" was arranged with the owner to encourage the dog to search for the reward. To meet the success criterion, dogs were required to approach the bowl and eat the food on at least six out of the seven "invisible" baiting trials. For the toy group, the owner threw the dog's favourite toy (which later served as the reward) two metres away and used a salient pre- arranged phrase to encourage the dog to search and retrieve it. In all the trials, the experimenter stood motionless beside the owner. To meet the success criterion, dogs were required to approach the toy and make appropriate contact with it (*i.e.,* sniff, pick up, paw the toy for at least 5 seconds, or bring it back to the owner), on at least seven out of ten trials. In both groups, dogs were reinforced with verbal praise after each trial.

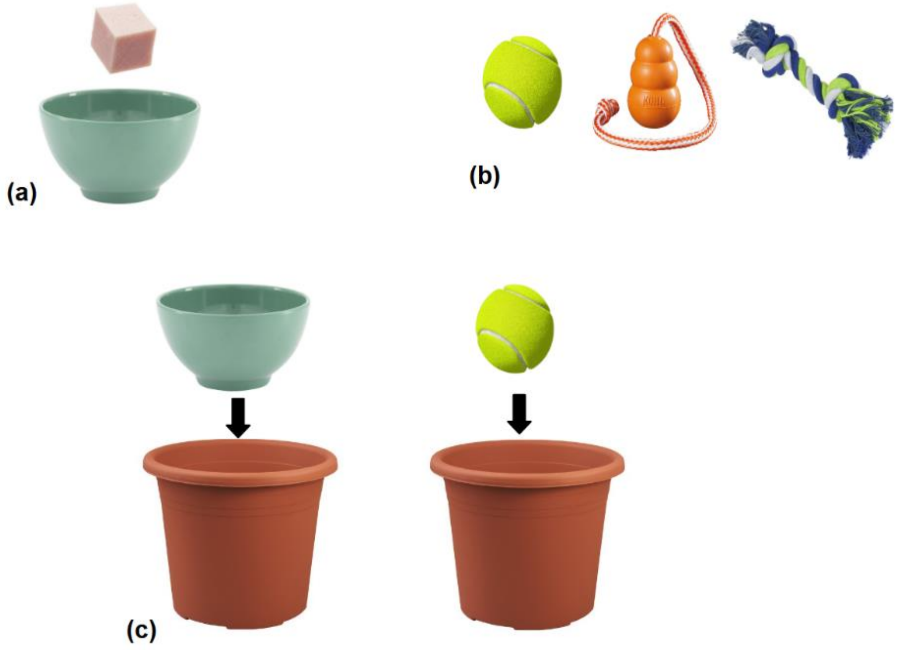

Figure 3: Materials (a) reward for the food group - one piece of the dog’s favourite food placed inside an opaque plastic container, (b) reward for the toy group - the dog’s favourite toy, (c) the displacement device – an opaque plastic flowerpot was used for both groups.

Introductory Trials of Single Visible Displacement: After screening, the dog and the owner entered the main testing room (see Figure 2). Dogs were given 1 minute to freely explore the room and the apparatus, while the experimenter explained the procedure and gave instructions to the owner. During this time, the experimenter and the owner did not interact with the dog. Subsequently, we ran six trials of *single visible displacement* where the experimenter visited one screen (clockwise for half of the subjects, and counterclockwise for the rest), and left the reward behind it. The order of screen visits was randomized, and each screen was visited twice. Each trial started with the owner sitting on a chair at the starting point, with the dog held in front of them, facing forward. If needed, owners manually held their dogs in position by the collar or shoulders to prevent them from initiating any movement ahead of time. Standing next to the owner, the experimenter captured the dog's attention with the reward in hand and placed it in the displacement device. The experimenter then walked to one of the screens, stood beside it, took the reward out of the container, and held it up for a moment before placing it back in the container. Then, she crouched down behind the screen, placed the object on the floor, and stood up after 3s. Standing next to the screen, the experimenter showed that the container was now empty. She then returned beside the owner and held the empty container by her side (at the height of the dog’s eyes), for the dog to inspect. The owner then gave the previously agreed command to encourage the dog to search for the reward. To eliminate any inadvertent cues, while the dog was searching the owner kept their eyes closed and the experimenter stood motionless gazing downwards. Dogs were allowed to search behind one screen only. After having made a choice – correct or incorrect - they were called back by their owner and were reinforced with verbal praise.

Success criterion: To meet the success criterion, dogs were required to be successful in at least five out of the six trials. A trial was considered successful if the dog approached the correct screen (*i.e.,* the screen with the hidden reward) on their first choice. If a dog visited the displacement device before approaching the correct screen, their search was still considered successful. Moreover, a trail was considered successful if the dog visited the correct screen but did not eat the food or make appropriate contact with the toy (*i.e.,* sniff, pick up, paw, or retrieve). A trail was considered *unsuccessful* if the dog first approached one of the irrelevant screens (*i.e.,* a screen without the hidden reward) or made no choice within 1 minute of release.

At this point, it is important to mention that the procedure of the introductory trials differed from the original Watson study in two significant ways: in the original study, introductory trials were *successive visible displacements,* and the reward was left behind one of the screens only in half of the trials (1st, 3rd and 5t - once per screen). In the other half of the trials (2nd, 4th, and 6th), the toy remained in the container. In contrast, in our case, introductory trials were *single visible displacements*, and the reward was left behind a screen in each trial. The hiding procedure was identical to Watson's trials 1, 3, and 5. The difference between *successive* and *single displacements* is that in the former, the experimenter visits all the screens, and leaves the object behind one of them, while in the latter the experimenter only visits one screen and leaves the object behind it. In both cases, the content of the container is revealed after a screen visit.

We decided to make these changes based on observations from pilot trials with the original procedure. Specifically, we noticed that the dogs we tested in the pilots behaved differently from those of the original study. As reported by (Watson et al., 2001), in Trials 1, 3, and 5, all dogs searched until they eventually found the reward, and all completed the task within 15 seconds. On Trials 2, 4, and 6, only 2 of the 19 dogs searched behind the screens, even after having seen that the object was still in the container. In our pilots, in Trials 1, 3, and 5 (toy left behind a screen), none of the dogs searched the screen containing the object on their first choice. Instead, many of the dogs wandered off to other locations in the room, and hence, did not complete the task. In Trials 2, 4, and 6 (toy remains in the container), most of the dogs searched behind the screens – despite having seen that the object was still in the container. Based on these observations, we decided to switch from *successive visible displacements* to *single visible displacements.* The purpose of these *simpler* trials was to inform dogs of the hiding potential of the three screens and the displacement device, as well as to ensure that dogs were motivated to search for the reward behind the screens. Pilots with the modified procedure indicated that dogs performed significantly better - *i.e.,* their first choice was the screen containing the object, they wondered off less, and, hence, were able to complete the task successfully.

Test Trial: In line with (Watson et al., 2001), the introductory trials were followed by one test trial of *successive invisible displacement*. This is a modified version of the standard successive invisible displacements task. The procedure was similar to the introductory trials of *single visible displacement*; however, this time the experimenter visited all three screens sequentially (clockwise for half of the subjects, and counterclockwise for the rest), and the displacements of the reward were *invisible* (*i.e.,* the content of the container was not revealed in between visiting the screens).

The trial started with the owner sitting on a chair at the starting point, with the dog held in front of them, facing forward. Standing next to the owner, the experimenter captured the dog's attention with the reward in hand and placed it in the displacement device. The experimenter then walked to the first screen, stood beside it, took the reward out of the container, and held it up for a moment before placing it back in the container. Then she placed the object back in the container and crouched down behind the screen. After 3s she got up and continued to the next screen.

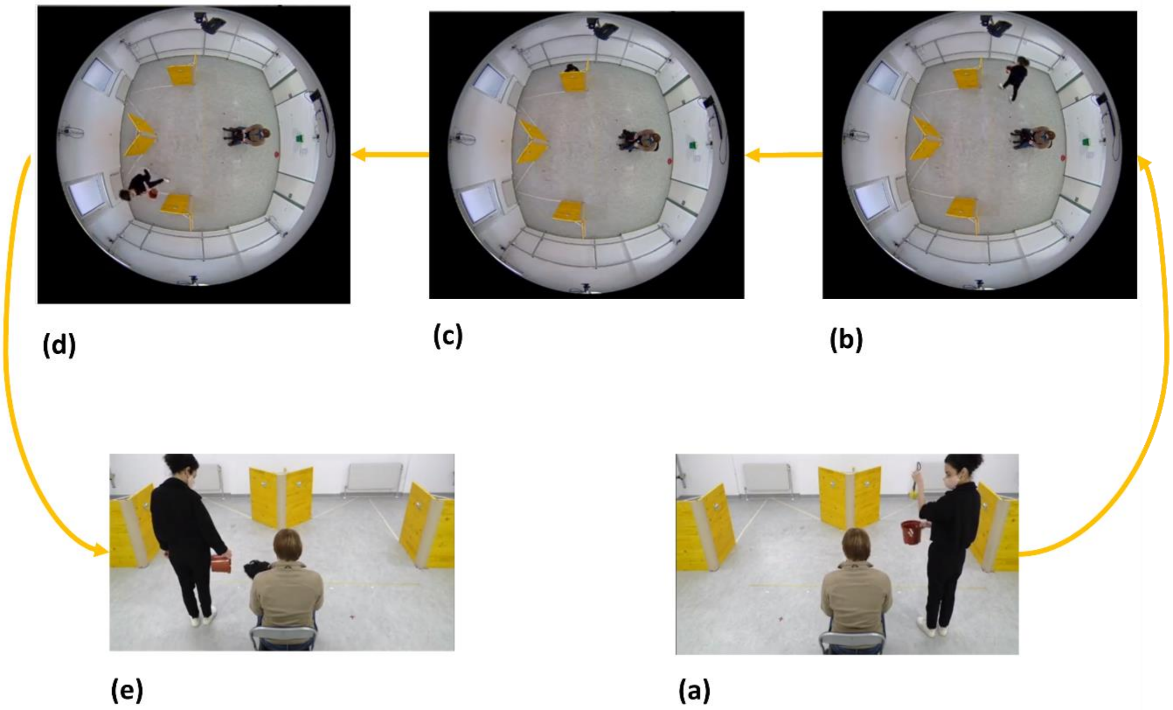

Figure 4: Overview of the test trial procedure: **(a)** dog and owner are seated at the starting point, the experimenter (the right side of the owner) shows the reward (toy) to the dog, **(b)** the experimenter places the reward back in the container and walks counterclockwise behind the screens, **(c)** experimenter crouches down behind a screen and serendipitously hides the reward in her pocket, **(d)** experimenter walks counterclockwise behind the screen to return to the starting point, **(e)** experimenter (left side of the owner) demonstrates the empty container to the dog who is then released for the search.

The same procedure was repeated for all the screens. The reward was not left behind any of the screens – instead, the experimenter placed it surreptitiously under her pullover, while crouching down behind one of the screens. After visiting all three screens, the experimenter returned beside the owner, and held the empty container by her side (at the height of the dog's eyes), for the dog to inspect. The owner then gave the previously agreed command to encourage the dog to search for the reward. The dog was given 1 minute to search or engage in any other type of behaviour. During this time, the owner kept their eyes closed and the experimenter stood motionless gazing downwards. At the end of the 1 minute, the hidden reward was presented to the dog by the experimenter. For photos of the procedure, see Figure 4.

Success criterion: The test trial was considered *successful* if the dog approached and explored each of the three (empty) screens at least once (in any order) within the 1-minute search period. If a dog visited the displacement device before approaching the screens, their search was still scored as correct. A trial was scored as *unsuccessful* if the dog made no choice or did not visit all three screens within 1 minute of release. *Unsuccessful* trials were excluded from further analysis.

2.5. Behaviour Coding

For subjects that passed the success criteria of the introductory and test trials, we coded four behavioural measures from the test trial: **(1)** *search trajectory* (direct vs. indirect), **(2)** *exploration time* for each screen (E1, E2, E3), **(3)** *travel time* taken to approach each screen (T1, T2, T3), and **(4)** *travel distance* covered while approaching each screen (D1, D2, D3). These 2 later measures we then used to calculate *pace* while approaching each screen (P1, P2, P3).

Search trajectory: A search trajectory was classified as *direct* if the subject progressed sequentially through the three screens, starting either from screen 1 or screen 3 and moving along a semi-circle without returning to a previously searched screen (Watson et al. 2001 refers to this type of search as *systematic)*. In contrast, a search trajectory was classified as *indirect* if it included re-exploration(s) of previously explored screens, passing by previously explored screen(s), visit(s) to the owner/experimenter, or additional travel(s) around the room. Search trajectories that began from the middle screen (as opposed to the first or last screen) were also classified as *indirect*.

Exploration time: Three exploration time values (E1, E2, E3 measured in seconds) were coded for each subject - one for each screen approached and explored. The act of exploration was defined as the dog standing or passing by pre-defined areas behind, or in front of each screen, with their head oriented toward the screen (visual exploration), or nose in physical contact/proximity with the screen or the surrounding area (olfactory exploration). For a detailed description of exploration behaviours, refer to the ethogram in Appendix C-1.

A re-exploration was coded if a dog explored a screen more than once. Re-explorations affected the travel time and travel distance measures - and eventually the pace calculation, which was based on these two values. Travel times included the time of any intermediate re-exploration trips (but did not include the time dedicated to re-exploring those screens). Likewise, travel distances included the distance covered during any intermediate re-exploration trips (but not the distance covered while re-exploring the screens). For more details, refer to Appendix C-2.

Travel time: Three *travel time* values (T1, T2, T3 measured in seconds) were coded for each subject – one for each screen approached and explored. T1 was the time that elapsed from the moment the dog started to move after receiving the search command (in the starting point) until the dog started exploring the first screen. T2 and T3, respectively, were the time that elapsed between the second and third screen explorations. The start of travel time was marked by the end of the previous exploration (*i.e.,* none of the exploration criteria were fulfilled) and the end was marked by the start of the next exploration. As mentioned above, if a travel involved re-explorations, travel time included the time taken to travel to the re-explored screens, but not the time spent exploring the screens. For further details about coding, refer to Appendix C-3.

Travel distance: Three *travel distance* values (D1, D2, D3) were coded for each subject - one for each screen approached and explored. D1 was the distance covered by the dog from the moment they started to move after receiving the search command (in the starting point) until the dog started exploring the first screen. T2 and T3, respectively, were the distance they covered between the second and third screen explorations. For further details about coding, refer to Appendix C-3.

A two-step procedure was followed for coding travel distance values. At first, we manually calculated travel distance values (measured in cm). We refer to this process as *manual* since the distance was calculated using the fixed distances between the starting point and the screens. For dogs that followed *direct* search trajectories *i.e.,* moved from the starting point to one of the outside screens (Screen 1 or 3) and then moved sequentially to the other screens along a semicircle, it was possible to calculate the exact distance covered in each travel (since the distances were known). However, this was not possible for dogs that followed *indirect* search trajectories (*e.g.,* did not move along the semicircle path, visited a location that was not one of the screens, re-explored a screen, etc.). To address this significant challenge, we turned to the use of K9-Blyzer (Canine Behaviour Analyzer), a powerful tool for automatic video analysis of canine behaviour which has already been used for a variety of scientific projects (see (Bleuer-Elsner et al., 2019; Fux et al., 2021; Karl et al., 2020; Menaker et al., 2022; Zamansky et al., 2021a, 2019)). K9-Blyzer's architecture (see Figure 5) consists of two modules: (a) a *tracking module* that uses object detection (in our context the object is the dog) to extract time-series data based on neural networks, and (b) a *feature analysis module*, which identifies, and measures the behavioural parameters of interest (in our context distance covered by the dog) based on the spatiotemporal data (trajectory) obtained from the tracking module.

The *tracking module* takes as input video footages of a dog freely moving in a room, potentially interacting with objects, humans, or other animals in the scene, along with a specified tracking method (*machine learning model*). The output is a time-series *tracking data* (represented in a JSON file) with the locations of detected objects in each video frame and measurements of specific parameters defined by the user to quantify behavioural parameters. Both the *tracking method* (*i.e.,* models used for detection), and the *scene* (including the amount of moving and fixed objects) can be easily adapted to fit the needs of specific experiments.

In the K9-Blyzer configuration used in this study, the *tracking method* was a neural network based on the Faster R-CNN architecture (Ren et al., 2017) which was specifically trained on dog images to enhance dog detection accuracy. The input *videos* were segments of the test trials showing a dog moving either from the starting point to a screen, or between two screens. The *scene* included three static objects (the three screens), and one dynamic object (the dog) detected by the *tracking method*. For each video, the output was the trajectory of the dog (for examples see Figure 6 c-f), and the distance covered by the dog in each video.

To calculate the distance travelled by the dog, Blyzer uses the *tracking method* to determine the dog's location in each video frame, generating bounding boxes around the dog (for examples see Figure 6 a-b). The center of a bounding box represents the dog's center of mass. Bounding boxes were saved in a JSON file representing a trajectory (*tracking data*). Post-processing operations were then applied to the tracking data to remove noise and enhance detection quality using smoothing and extrapolation techniques, achieving almost perfect detection rates (averaging above 98% on our dataset). Subsequently, the center of mass of the dog was computed in each frame, and the distance in pixels between consecutive frames was calculated. By aggregating these distances over the entire video, we obtained the dog's trajectory and calculated the total distance covered by the dog during the observation period (a single travel event). Hence, for each subject, we obtained three *automatic* travel distance values (D1, D2, D3 measured in pixels).

Pace: Three *pace* values (P1, P2, P3) were calculated for each subject - one for each screen approached and explored. Pace was defined as travel time (P1, P2, P3) divided by distance (D1, D2, D3). For dogs that followed indirect search trajectories involving re-explorations, total travel time was calculated by summing up all intermediate travel times, and total travel distance was calculated by summing up all intermediate travel distances. Manual pace values (measured in sec/cm) were calculated using the manual travel distance values. Automatic pace values (measured in sec/pixel) were calculated using the travel distance values calculated by K9-Blyzer. Subsequent statistical analysis was carried out using the automatic pace values. For further details about coding, refer to Appendix C-4.

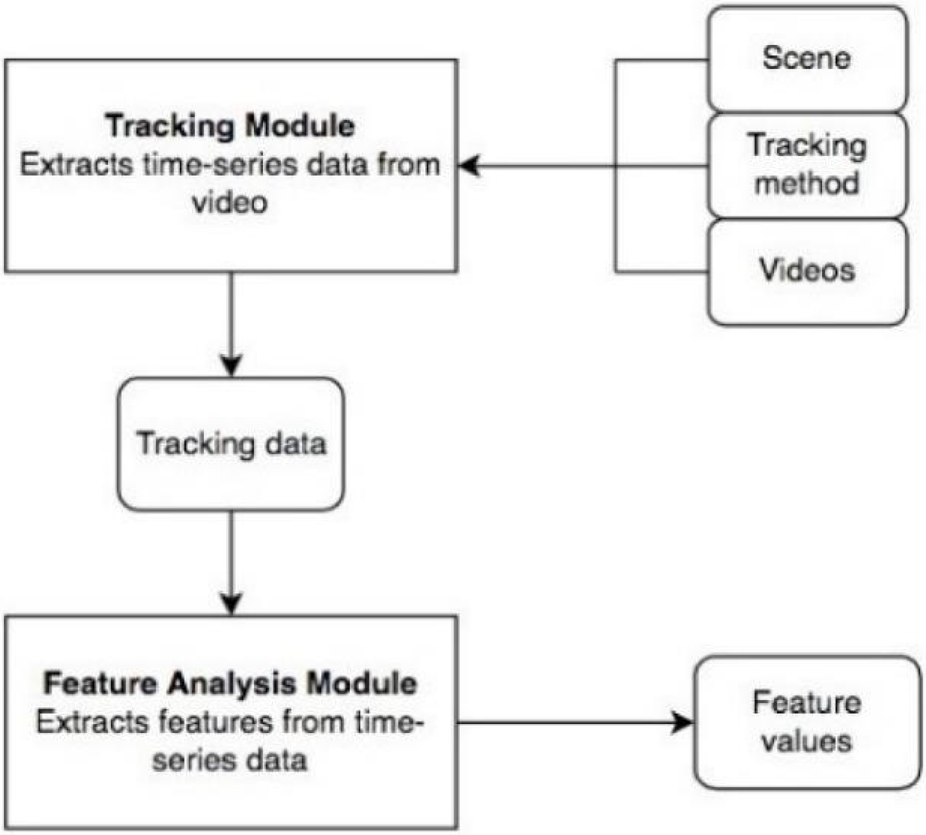

Figure 5: The architecture of K9-Blyzer, including the two main modules (tracking module, feature analysis module) along with their input and output.

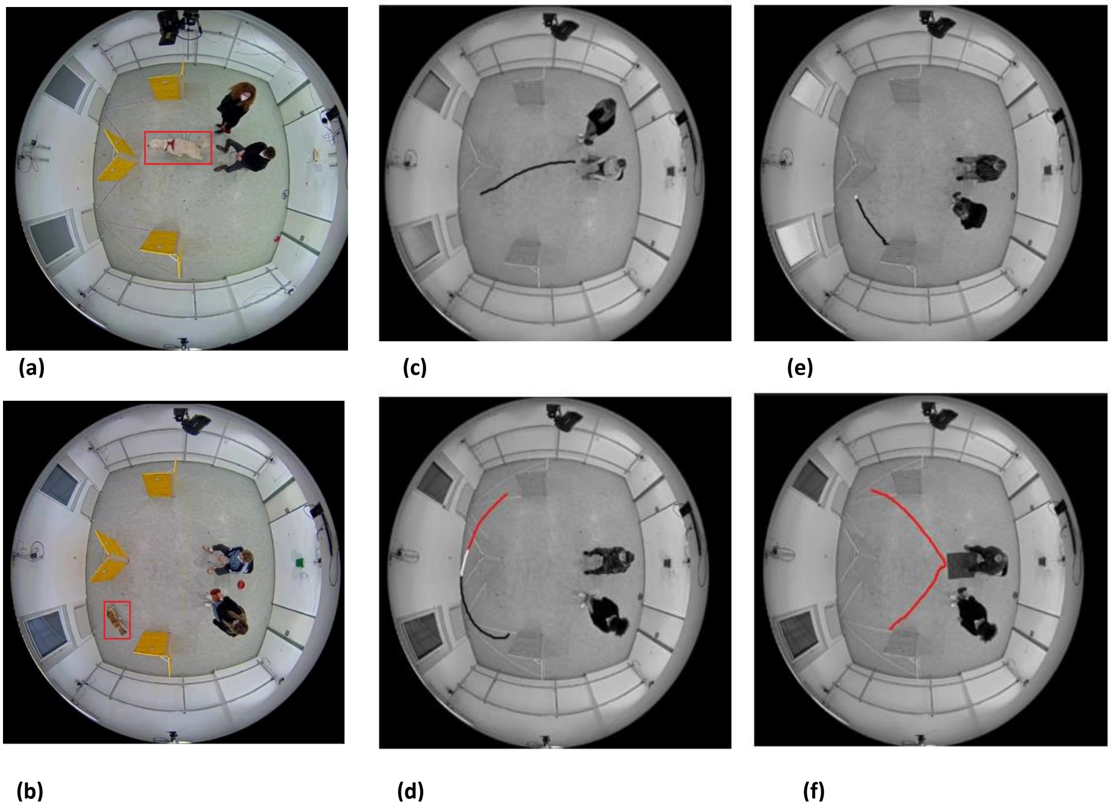

Figure 6: Examples of bounding boxes used to determine dogs' location in each video frame (a,b) and dog travel trajectories (c-f). The center of the bounding box marks the center of mass of the dog in a frame. The dog's trajectory is calculated by concatenating the center of mass over consecutive frames. Examples of search trajectories: (c) direct travel from the starting point to screen 2, (d) indirect travel from screen to screen involving a pass behind the middle screen, (e) direct travel from screen 1 to screen 2, (f) indirect travel from screen to screen where the dog wandered off to visit the owner in between screen visits.

2.6. Interrater reliability

Exploration times and travel times were coded by one of the experimenters. A randomly selected subset of the trials (93 out of a total of 129 observations) was re-scored by a second experimenter. We calculated an Intra-class Correlation Coefficient (ICC) (McGraw 1996) on this subset of exploration time observations to assess the agreement between two coders. The resulting inter-rater reliability was good (ICC: 83.5%; N=93, p<0.001).

To assess the agreement between the manual and automatic pace values, we calculated an ICC on a subset of observations (48 out of a total of 129 observations). Only trials of directly searching dogs were included in this analysis. As manual and automatic pace values were calculated on different metric scales (sec/*cm* for manual pace, sec/*pixels* for automatic pace, mean+-sd manual pace: 1.417+-1.259, automatic pace: 0.022+-0.011) we calculated ICC on the scaled values (0-1). The resulting inter-rater reliability was good (ICC: 83.7%; N=48, p<0.001).

2.7. Statistical Analysis

We fitted two separate Linear Mixed Models (Baayen, 2008) using maximum likelihood estimation in R (version 4.2.3; R Core Team, 2022) with the 'lmer' function from the 'lme4' package (Bates et al., 2015). The models had *pace* and *exploration time* as response variables. To meet the assumptions of linear models, both response variables were log-transformed (base e). Gaussian error structures with an identity link function were employed in the models. The fixed effects part for each model was identical, with the primary factors of interest being *reward* (with levels Food and Toy), *screenNumber* (ranging from 1 to 3 - representing the three screens), and their interaction. Additionally*, experimenter* (with levels CT, KB, and MAR) served as a control predictor. The term *screenNumber* was z-transformed to facilitate model convergence and enhance the interpretability of model estimates (Schielzeth, 2010). Detailed information on the models can be found in Appendix D1.

To account for repeated observations of the same individual, random intercept effects of subject were included in the models. Moreover, to control for the Type I error rate at the nominal level of 0.05 (Barr et al., 2013; Schielzeth & Forstmeier, 2009), we included the random slope of *screenNumber* within subject. However, the parameter for the correlation between the random intercept and slope was removed from the exploration time model since it was estimated to be essentially 1, rendering it unidentifiable (Matuschek et al., 2017).

To address the increased Type I error risk due to multiple testing (Forstmeier & Schielzeth, 2011), we initially tested the overall effect of the fixed effects of interest. Likelihood ratio tests were employed to compare the full model with all terms included to a null model lacking the effects of *reward* and *screenNumber* (Dobson & Barnett, 2018). If the full-null comparison showed significance, we proceeded to test individual fixed effects using the Satterthwaite approximation (Luke, 2017) with the 'lmerTest' package (version 3.1-3) (Kuznetsova et al., 2017), fitting models with restricted maximum likelihood. Model confidence intervals were estimated using 1,000 parametric bootstraps via the 'bootMer' function from the 'lme4' package.

After fitting the models, we ensured that none of the model assumptions were violated and assessed model stability. Residual plots and qq-plots showed no substantial deviations from the assumptions of normality and homogeneity of residuals for each model. The Variance Inflation Factor (VIF), calculated using the 'car' package (version 3.0-12) (Fox & Weisberg, 2019), indicated no issues with collinearity (max VIF = 1). Additionally, we visually confirmed that the best linear unbiased predictors (BLUPs) per level of the random effects were approximately normally distributed (Baayen, 2008). To assess model stability (Nieuwenhuis et al., 2012), we systematically excluded each level of random effects and compared the resulting estimates to those obtained from the model based on all data, demonstrating good stability.

2.8. Results & Discussion

Screening & Introductory trials

In total, 56 subjects were tested (food group N= 29, toy group N =27). While all subjects passed the screening trials, 13 failed to meet the success criterion in the introductory visible displacement trials and were excluded from further analysis. All the remaining 43 subjects (food group N=22, toy group N=21) passed the success criterion in the test trial and were included in the final analysis. For a detailed breakdown of the numbers, refer to Appendix B.

Test trial

The full model for the response variable *pace* provided a significantly better fit to the data compared to the respective null model ($\chi^2 = 7.99$, df = 3, p = 0.046) – for details refer to Appendix D1. Inspecting the test predictors did not reveal a significant interaction between *reward* and *screenNumber* (p=0.90). However, we found a significant effect of *screenNumber* (p=0.006) suggesting that individuals increased their pace as the search progressed. For full and reduced model outputs see Table 1 (a) and (b) respectively.

Concerning *exploration time*, the analysis revealed no significant difference between the full model and the respective null model ($\chi^2 = 5.75$, df = 3, p = 0.125). For full model output see Table 1 (c). For a graphical representation of the *pace* and *exploration values* in Experiment 1 refer to Figure 7.

The results of Experiment 1 confirm the presence of the *slowing down effect* reported in the original study of Watson et al. Specifically, the significant positive relationship between *screenNumber* and *pace* indicates that as dogs advanced through the search, their pace consistently increased (*i.e.,* they moved slower). Exploration time remained constant across screen explorations (as indicated by the non-significant full-null model comparison). In conjunction, these findings reveal that dogs slowed down precisely because they took longer to travel between screens, and not because they spent more time exploring the screens. The analysis did not reveal a significant interaction between *reward* and *screenNumber* – neither in terms of pace nor in terms of exploration time. Hence, the type of reward did not influence the behaviour of dogs.

One notable observation is the fact that, in comparison with the values reported in the original study, the proportion of dogs that followed an *indirect* search trajectory in the test trial was particularly high (original study: direct search 69%, indirect search 31%; experiment 1: direct search 51%, indirect search 49% - for details see Appendix B). Indirect trajectories included re-explorations of previously explored screens, visits to the owner/experimenter between screen visits, additional travel around the room, or a search starting from the middle screen – which led to longer travel distance.

Several reasons could account for the high number of indirect searching dogs in our study (see Chapter 4 for a detailed discussion). The main concern related to this unexpected behaviour was the fact that it introduced noise to the data, likely making it difficult to detect the predicted effects. To address this concern, we designed a follow-up experiment (Experiment 2 – The updated Watson task) where, on the one hand we tried to minimize the number of indirectly searching dogs, and on the other hand, we controlled for this variation in the statistical analysis.

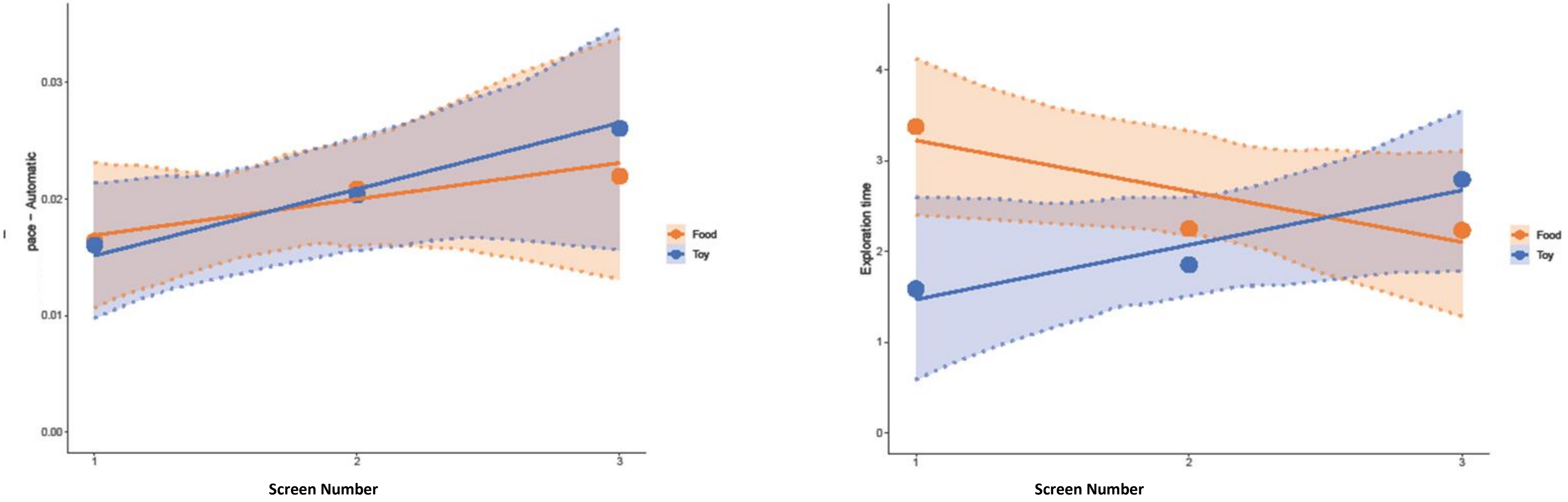

Figure 7: Comparison of pace (*left*) and exploration time (*right*) while approaching the first, second the third screen. Values are plotted as a function of *reward* and *screenNumber*, with the factor experimenter centered to a mean of zero. The dots show the average value, solid lines show the fitted model, and dashed lines the 95% intervals.

Table 1: (a) Full model output for the response variable pace, (b) Reduced model output for the response variable pace, (c) Full model output for the response variable exploration time.

(a) Pace – Results of Full Model (Experiment 1)

Term	**Estimate**	***SE***	**lower cl**	**upper cl**	df_1	df_2	***F*-value**	***p*-value**	**min**	**max**
(Intercept)	-3,96	0,11	-4,17	-3,75				(3)	-4,00	-3,90
*reward*Toy(1)	0,01	0,11	-0,19	0,23				(3)	-0,04	0,05
z.*screenNumber*(2)	0,13	0,06	0,00	0,25				(3)	0,12	0,14
experimenterKB	0,13	0,13	-0,11	0,38				(3)	0,04	0,17
experimenterMAR	-0,49	0,12	-0,71	-0,25	2,00	39,05	12,83	0,00	-0,53	-0,42
*reward*Toy:z.*screenNumber*	-0,01	0,09	-0,18	0,16	1,00	41,07	0,02	0,90	-0,06	0,03

(b) Pace – Results of Reduced Model (Experiment 1)

Term	**Estimate**	**SE**	**lower cl**	**upper cl**	**df1**	**df2**	**F-value**	**p-value**	**min**	**max**
(Intercept)	-3,96	0,10	-4,17	-3,76				(3)	-4,01	-3,90
*reward*Toy	0,02	0,11	-0,19	0,22	1,00	39,05	0,02	0,89	-0,03	0,05
z.*screenNumber*	0,13	0,04	0,04	0,21	1,00	42,07	8,48	0,01	0,10	0,15
experimenterKB	0,13	0,13	-0,13	0,37				(3)	0,04	0,17
experimenterMAR	-0,49	0,12	-0,72	-0,25	2,00	39,01	12,86	0,00	-0,53	-0,42

(c) Exploration time – Results of Full Model (Experiment 1)

Term	**Estimate**	**SE**	**lower cl**	**upper cl**	**df1**	**df2**	**F-value**	**p-value**	**min**	**max**
(Intercept)	0,36	0,17	0,03	0,70				(3)	0,27	0,49
*reward*Toy	-0,22	0,18	-0,57	0,13				(3)	-0,31	-0,13
z.*screenNumber*	-0,14	0,12	-0,39	0,09				(3)	-0,20	-0,10
experimenterKB	-0,05	0,22	-0,48	0,38				(3)	-0,17	0,08
experimenterMAR	0,59	0,20	0,23	0,98	2,00	38,93	5,70	0,01	0,52	0,65
*reward*Toy:z.*screenNumber*	0,34	0,17	0,01	0,69	1,00	41,00	4,10	0,05	0,28	0,41

(1) *reward* and experimenter were dummy coded with 'food' and 'CT; being the reference categories respectively.
(2) *screenNumber* was z-transformed to a mean of 0 and a standard deviation of 1, mean (sd) was 1.98 (0.82)
(3) not indicated because of having a very limited interpretation

3. Experiment 2 – The Updated Watson Task

The *Updated Watson Task* is a modified version of Experiment 1, where the experimental procedure and the statistical models are refined to account for the *indirect* search behaviour of dogs in the test trial. In terms of the experimental procedure, six additional introductory trials of visible displacement were introduced to encourage dogs to search directly in the test trial. In contrast to the introductory trials of Experiment 1, where dogs saw the experimenter walk behind one screen and leave the object there, in these trials the experimenter walked sequentially behind all the screens and left the object behind one of them. The contents of the container were revealed after each screen visit. We expected that the additional trials would: (a) provide dogs with more information about the spatial layout of the area to be searched, and (b) demonstrate that the most effective way to search is to walk behind the screens sequentially (*i.e.,* following a direct trajectory). In terms of the statistical analysis, the term *searchBehaviour* (with levels direct and indirect) was included as an additional control predictor in the models for *pace* and *exploration time*.

3.1. Subjects

Dogs were recruited and assigned to the toy and food groups following the same procedure as in Experiment 1. In total, 72 dogs naïve to the experimental procedure were recruited (35 males, 37 females, age: mean = 5,86, SD =3,50 - for demographic details see Appendix A); 42 dogs were assigned to the food group (13 males, 26 females, age: mean = 6,79, SD =3,39) and 33 dogs were assigned to the toy group (22 males, 11 females, age: mean = 4,76, SD =3,30). We counterbalanced the groups as much as possible for age, sex, and breed.

3.2. Experimental Setup

The experimental setup and apparatus were the same as in Experiment 1 (see Section 2.3). The only difference was that for the food group, the *green* opaque bowl that held the piece of food was replaced with a *white* opaque bowl. Since the experimenter was wearing a black pullover throughout the session, a white bowl was more visible than a green one (due to the high contrast with the experimenter's black pullover).

3.3. Experimental Procedure

The test was administered within a single session lasting approximately 30 min on average. Dogs were tested individually with their owner present. Two experimenters ran the tests. The experimental procedure is outlined below.

Screening Trials: Procedure and success criterion identical to those of Experiment 1.

Introductory Trials of Single Visible Displacement: Procedure and success criterion identical to those of Experiment 1.

Introductory Trials of Successive Visible Displacement: We run six trials of Successive Visible Displacement. The procedure was identical to the single visible displacement trials, however, this time the experimenter visited all the screens (sequentially walking on a semi-circle route clockwise for half of the subjects, and counterclockwise for the rest), and after each screen visit, showed whether the reward was still in the container or not. The reward was left behind one of the screens (twice behind each screen, in a randomized order). Each trial started with the owner sitting on a chair at the starting point, with the dog held in front of them, facing forwards and the experimenter standing next to the owner with the container in hand. The experimenter walked to the first screen, stood beside it, took the reward out of the container, and held it up for a moment before placing it back in the container. Then, she crouched down behind the screen, placed the object on the floor (or not), and stood up after 3s. Standing next to the screen, the experimenter showed that the container was now empty, or took out the object, and held it up for a moment, before placing it back in the container. After repeating this process for all three screens, the experimenter returned beside the owner, and held the empty container by her side (at the height of the dog's eyes), for the dog to inspect. The owner then gave the previously agreed command to encourage the dog to search for the reward. The dogs were allowed to search until they found the reward – there was no time restriction – and they could visit more than one screen. After having made a choice – correct or incorrect- they were called back by their owner and were reinforced with verbal praise.

Success criterion: To meet the success criterion, dogs were required to be successful in at least five out of the six trials. A trial was considered successful if the dog approached the correct screen (*i.e.,* the screen with the hidden reward) on their first choice. There was no time limitation. If a dog visited the displacement device before approaching the correct screen, their search was still considered successful. Moreover, a trial was considered successful if the dog visited the correct screen but did not eat the food or make appropriate contact with the toy (*i.e.,* sniff, pick up, paw, or retrieve). A trial was considered *unsuccessful* if the dog first approached one of the empty screens (*i.e.,* a screen without the hidden reward).

Test Trial: Procedure and success criterion identical to those of Experiment 1.

3.4. Behavioural Coding

The videos were coded by one of the experimenters following the same procedure as in Experiment 1. For subjects that passed the success criteria of the introductory and test trials we coded four behavioural measures from the test trial: **(1)** *search trajectory* (direct vs. indirect), **(2)** *exploration time* for each screen (E1, E2, E3), **(3)** *travel time* taken to approach each screen (T1, T2, T3), and **(4)** *travel distance* covered while approaching each screen (D1, D2, D3). Based on the two latter values, *pace* was then calculated while approaching each screen (P1, P2, P3). Travel distance was calculated using K9-Blyzer, and those values were subsequently used to calculate pace which was used for the statistical analysis.

3.5. Statistical Analysis

We fitted two Linear Mixed Models (Baayen, 2008) with *pace* and *exploration time* as the response variables (for details see Appendix D2). To meet the assumptions of linear models, both response variables were log-transformed (base e). Gaussian error structures with an identity link function were employed in the models. The models were identical to those of Experiment 1, with the distinction that *searchBehaviour* (with levels direct and indirect) was included as an additional control predictor to account for the indirect search behaviour of dogs observed in Experiment 1. The analysis was identical to that described in Experiment 1.

3.6. Results & Discussion

Screening & Introductory trials

In total, 72 subjects were tested (food group N=39, toy group N=33). From those, 3 subjects failed to meet the success criterion in the screening trials, and 17 failed the introductory trials of visible displacement - after excluding those subjects, the sample size was down to 55 subjects. Of those, all passed the introductory trials of invisible displacement, but 10 failed the success criterion in the test trial. The final sample consisted of 45 subjects (food group 24, toy group 21). For a detailed breakdown of the numbers, refer to Appendix B.

Test Trial

In agreement with the results of Experiment 1, the full model for the response variable *pace* provided a significantly better fit to the data compared to the respective null model ($\chi^2 = 22.71$, df = 3, $p < 0.001$) - for details refer to Appendix D2. Inspecting the test predictors did not reveal a significant interaction between *reward* and *screenNumber* ($p = 0.338$). However, we found a significant effect of *screenNumber* ($p < 0.001$) suggesting that individuals increased their pace as the search progressed. For model outputs see Table 2 (a) and (b).

For *exploration time*, the analysis revealed no significant difference between the full model and the respective null model ($\chi^2 = 4.47$, df = 3, $p = 0.215$). For full model output see Table 2 (c). For a graphical representation of the *pace* and *exploration values* refer to Figure 8.

The six introductory trials of invisible displacement which were included in the experimental procedure did not encourage dogs to search directly in the test trial. We expected that after repeatedly witnessing the experiment demonstrate a direct route behind the screens, dogs would either learn that a direct search is the most effective way to search, or they would be more inclined to mimic the route of the experimenter, and thus search directly. Nevertheless, the proportion of dogs that searched *indirectly* remained as high as in Experiment 1 (Experiment 1: direct search 51%, indirect search 49%; Experiment 2: direct search 56%, indirect search 44% - see Appendix B). To control for the effects of the variation in search trajectories*, searchBehaviour* was added in the models of pace and exploration time. The results of Experiment 2 are fully consistent with those of Experiment 1 and reinforce the presence of the *slowing down effect*. Specifically, the significant effect of *screenNumber* indicates that as dogs advanced through the search, their pace consistently increased (*i.e.,* they moved slower). In line with Experiment 1, exploration time remained constant across screen explorations (as indicated by the non-significant full-null model comparison). Even after controlling for the variation in dogs' search behaviour, we did not find an effect of reward type on the dogs' pace, or exploration time.

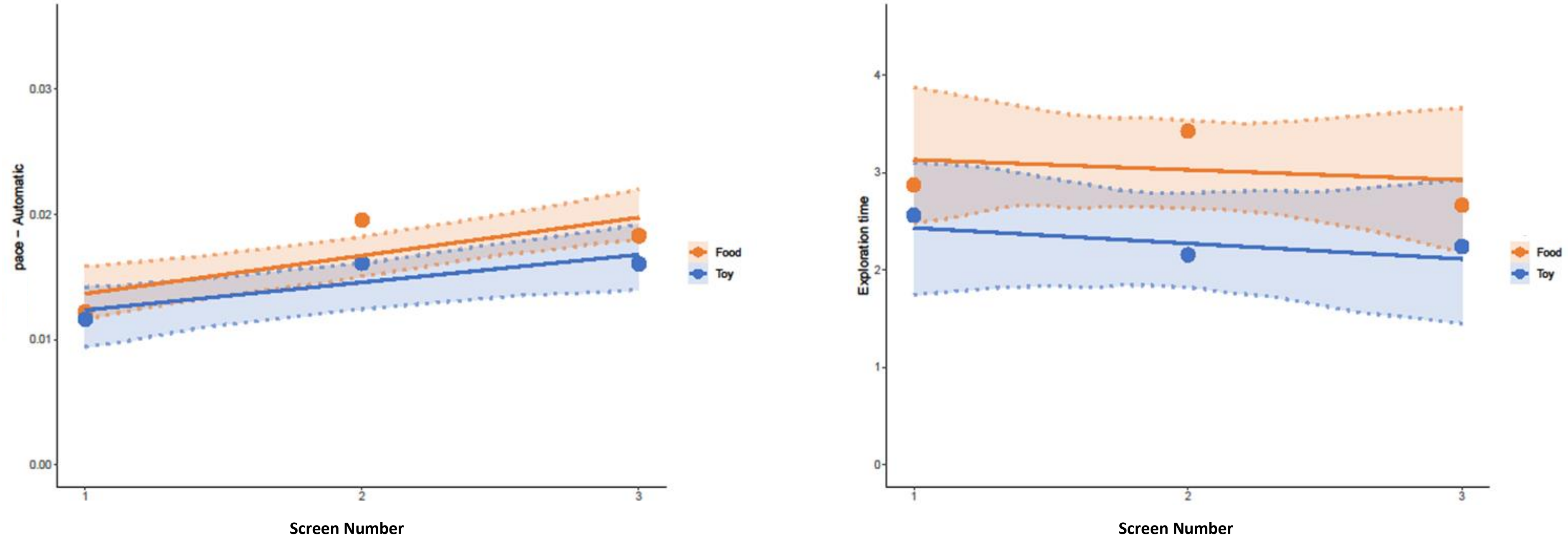

Figure 8: Comparison of pace (*left*) and exploration time (*right*) while approaching the first, second the third screen. Values are plotted as a function of *reward* and *screenNumber*, with the factor experimenter centered to a mean of zero. Dots show average value, solid lines the fitted model, and dashed lines the 95% intervals.

Table 2: (a) Full model output for the response variable pace, (b) Reduced model output for the response variable pace, (c) Full model output for the response variable exploration time.

(a) Pace – Results of Full Model (Experiment 2)

Term	Estimate	SE	lower cl	upper cl	df1	df2	F-value	p-value	min	max
(Intercept)	-4,17	0,07	-4,31	-4,02				(3)	-4,20	-4,14
*reward*Toy(1)	-0,12	0,08	-0,27	0,03				(3)	-0,14	-0,08
z.*screenNumber*(2)	0,16	0,04	0,08	0,23				(3)	0,14	0,17
searchBehaviourindirect	-0,02	0,08	-0,18	0,13	1,00	41,00	0,08	0,78	-0,06	0,00
experimenterMAR	0,01	0,08	-0,17	0,16	1,00	41,00	0,01	0,94	-0,02	0,03
*reward*Toy:z.*screenNumber*	-0,06	0,06	-0,16	0,06	1,00	88,00	0,93	0,34	-0,08	-0,04

(b) Pace – Results of Reduced Model (Experiment 2)

Term	Estimate	SE	lower cl	upper cl	df1	df2	F-value	p-value	min	max
(Intercept)	-4,17	0,07	-4,31	-4,02				(3)	-4,20	-4,14
*reward*Toy	-0,12	0,08	-0,28	0,02	1,00	41,00	2,38	0,13	-0,14	-0,08
z.*screenNumber*	0,13	0,03	0,08	0,19	1,00	89,00	21,72	0,00	0,12	0,14
searchBehaviourindirect	-0,02	0,08	-0,17	0,13		41,00	0,08	0,78	-0,06	0,00
experimenterMAR	0,01	0,08	-0,16	0,18	1,00	41,00	0,01	0,94	-0,02	0,03
*reward*Toy:z.*screenNumber*	-4,17	0,07	-4,31	-4,02				(3)	-4,20	-4,14

(c) Exploration time – Results of Full Model (Experiment 1)

Term	Estimate	SE	lower cl	upper cl	df1	df2	F-value	p-value	min	max
(Intercept)	0,79	0,14	0,53	1,05				(3)	0,74	0,84
*reward*Toy	-0,30	0,15	-0,59	-0,03				(3)	-0,36	-0,24
z.*screenNumber*	-0,01	0,07	-0,15	0,12				(3)	-0,04	0,01
searchBehaviourindirect	0,28	0,15	-0,01	0,56	1,00	41,00	3,61	0,06	0,24	0,34
experimenterMAR	-0,02	0,15	-0,32	0,29	1,00	41,00	0,02	0,88	-0,09	0,04
*reward*Toy:z.*screenNumber*	0,03	0,10	-0,16	0,23	1,00	43,00	0,09	0,77	-0,03	0,07

(1) *reward* and experimenter were dummy coded with ‘food’ and ‘LKL; being the reference categories respectively

(2) *screenNumber* was z-transformed to a mean of 0 and a standard deviation of 1, mean (sd) was 2 (0.82)

(3) not indicated because of having a very limited interpretation

4. General Discussion

What cognitive processes are involved in finding a ball that has rolled under the couch? Are the same processes involved when scavenging for a hidden treat? The central question addressed in this thesis was whether the nature of the reward, whether it be food or a toy, influences the search behaviour of dogs. The collective findings of the two experiments reported here (a) reaffirm the original conclusion put forth by (Watson et al., 2001) that dogs predominantly employ associative strategies when searching for a hidden reward, and (b) reveal that the reward's nature (be it food or toy) does not influence their search behaviour. These findings provide valuable insights into the problem-solving abilities of dogs in the physical domain and call for additional exploration in this area. Importantly, this work adds to the sparse body of robust methods that distinguish between the use of inference and association.

4.1. Effects of Reward Type

Building upon previous research suggesting that different types of rewards might trigger distinct cognitive strategies or elicit varied performance in dogs, we tested two alternative hypotheses for the effect of reward on the search behaviour of dogs. H1 posited that different rewards might trigger the use of different cognitive strategies, while H2 posited that dogs may represent rewards differently based on their inherent characteristics. For each hypothesis, we generated two predictions relating to search pace and exploration time.

Our empirical findings yielded no substantiation for either hypothesis, and none of our predictions was confirmed. Rather, the results revealed that the intrinsic nature of the reward (be it food or toy), on its own, does not significantly influence the search behaviour of dogs. In Experiment 1, we observed that while searching for a reward that was concealed inside a container and invisibly disappeared behind three screens, dogs slowed down as they approached the final screen, and they did so independently of whether the reward was food or a toy. After failing to find an effect of reward type on the pace or exploration time of dogs in Experiment 1, in Experiment 2, the analysis was extended by including search behaviour (direct vs. indirect) as an additional control predictor. Nevertheless, even after controlling for this, we found no evidence that the reward type would affect the search behaviour of dogs.

Our results contradict the two hypotheses proposed by Belger & Bräuer (2018), suggesting that dogs may employ distinct cognitive strategies or hold diverse mental representations for different reward types. Instead, it appears that the simpler explanation rooted in impulsivity that they proposed (*i.e.,* food-rewarded dogs may exhibit higher impulsivity and are therefore less likely to make correct choices or show metacognitive abilities) better aligns with the observed higher performance of toy-searching dogs in their study. Another study demonstrating significant performance disparities in toy vs. food-rewarded dogs was conducted by (Range et al., 2011). In contrast to (Belger & Bräuer, 2018), in this study, dogs performed significantly better in a means-end understanding task if the reward was food instead of a toy that could be exchanged for a food reward. Yet, once again, the differences in performance can be attributed to the stimulating value of the reward and the dogs' impulsivity. This interpretation is bolstered by the following observations: (a) the testing cohort only included dogs with high motivation for toy retrieval, (b) there were brief play sessions with the toy reward between trials, and (c) toy-rewarded dogs required fewer sessions to complete the task compared to food-rewarded dogs.

The abovementioned findings highlight the pivotal role of reward type and the possibility of inhibitory control or motivation-related issues in the study of dogs' physical problem-solving abilities. It could be argued that motivation may have influenced the behaviour of dogs in our experiments, potentially masking the effect of reward type. We employed a rigorous selection process, only including dogs in the toy/food groups who exhibited high motivation to retrieve the toy/consume the food during screening trials. Highly motivated dogs may have encountered challenges in conducting a thorough direct search of the screens or effectively using perceptual cues to deduce the location of the reward. However, we believe this scenario is improbable for two reasons: (a) the statistical analysis revealed no main effect of reward type either on pace or on exploration time; that is, there was no evidence that the type of reward would affect the motivation of dogs, and (b) the distribution of direct vs. indirect searching dogs was nearly identical in both the food and toy groups (for detailed statistics refer to Appendix B). Indirect searching could potentially signify a lower level of interest in the reward, as evidenced by the tendency of dogs to interrupt their search by visiting their owner or to casually explore the room, rather than concentrating on the hiding screens.

In conclusion, our findings challenge the previously held notion that reward type has a substantial influence on the cognitive strategies employed by dogs during object search in physical problem-solving tasks. Future research endeavours are needed to explore whether dogs form distinct mental representations of food and toy rewards or if these rewards elicit the utilization of divergent cognitive strategies.

4.2. Watson Study Replication

Our experiments serve as a replication of the original Watson study and contribute to the body of evidence challenging the use of inferential reasoning by dogs during object-search tasks. Rather, our results indicate that, regardless of the reward type, dogs' behaviour appears to be guided primarily by the association between the screens and the concealed reward. The observed *slowing down effect* suggests that the strength of this association wanes with each unsuccessful attempt to find the reward, consequently diminishing the dogs' anticipation of finding it behind subsequent screens.

In comparison to the original Watson study, we tested a significantly larger dog population (Original study N= 30, Experiment 1: N= 56, Experiment 2: N= 72,) and used a more rigorous analysis to scrutinize the presence of the *slowing down effect*. Unlike the compound measure of *speed* used by Watson et al. (which combined total time spent travelling between screens and exploring screens), we conducted a separate analysis of *pace* while travelling between screens and screen *exploration time*. Pace measurements were calculated with K9-Blyzer, a tool for automatic video analysis of canine behaviour (Amir et al., 2017). Traditionally, in animal-related disciplines, behavioural analysis is done through direct human observation and manual coding. However, this is a laborious and tedious task, prone to subjectivity and leaving ample room for error (for a detailed discussion of the limitations of manual coding see (Anderson & Perona, 2014). Automated tools that use artificial intelligence to analyse animal behaviour not only empower human observers in terms of accuracy and volumes of processed data but may also reveal new characteristics of behaviour which are inaccessible by direct human observation (Zamansky et al., 2021b). In our experiments, due to the diverse indirect trajectories that dogs used during their search, coding the pace measurements would not have been possible, without the use of K9-Blyzer.

The results of our experiments consistently and robustly demonstrate that, as dogs progressed in their search, (a) their pace significantly increased, while (b) their exploration time remained constant. Taken together, these two observations reveal that dogs slowed down because they took longer to travel between screens, and not because they dedicated more time to screen explorations. This distinction is important, since extended exploration behind later screens may signify either *surprise* at not finding the reward or a *persistent expectation* of finding the reward - both clear indicators of the use of inferential reasoning by exclusion (*i.e.,* inferring that if the object is not behind screen 1 and not behind screen 2, then it must be behind screen 3).

In contrast to our findings, earlier studies suggest that dogs use inferential reasoning when searching for a toy in a two-way cup task (Erdohegyi et al., 2007). If dogs have the capacity for inferential reasoning, why didn't they make use of it in the Watson task? One plausible explanation for the dogs' apparent failure to use inferential reasoning in the Watson task might stem from the inherent complexity of the task, which hinges on the concept of successive invisible displacement (object permanence Stage 6). (Piaget, 1952) argued that successful performance in the invisible displacement task is evidence of mental reconstruction of the reward's trajectory. Drawing inferences about the trajectory of a reward placed inside a moving container, and mentally representing its unseen hiding location, may be more challenging than retaining the mental representation of a reward concealed within a stationary container, as in the two-way cup task (Espinosa, 2022). Moreover, subjects could succeed in the two-way cup by simply avoiding the empty container, instead of inferring that the other cup is baited (Völter & Call, 2017).In contrast, in the Watson task, subjects do not witness any of the hiding screens being empty, making the avoidance explanation less likely.

In the study of (Erdohegyi et al., 2007), human social-communicative cues (*e.g.,* directional gestures, attention) prevented dogs from solving the two-way choice task using inferential reasoning. In our experiments, we eliminated all communicative cues from the owner and experimenter, so it is unlikely that the behaviour of dogs was influenced this way. However, given the complexity of the task, it is possible that dogs expected to receive a cue from their owner to find the reward, instead of using perceptual cues and their own reasoning abilities. This might also explain why, after failing to find the reward behind the first or second screen, some dogs visited their owner - presumably to ask for help. Other studies (*e.g.,* (Bräuer et al., 2006; Szetei et al., 2003)) also show that in situations when direct visual information about a hidden reward is missing, human communicative and behavioural cues indicating the hiding place is of great importance for dogs.

Perhaps, instead of representing the reward, and making inferences about its location based on visual cues, dogs used their keen sense of smell to track and locate the food and toys in our experiments (Horowitz & Franks, 2020). The behavioural coding analysis indicates that dogs often sniffed at the screens, and the area around them, however, it is unlikely that olfaction was their primary search modality since they always looked, or at least glanced, at the area behind the screens. Hence, it seems more plausible that dogs used their noses to gain additional information when they did not find the reward by vision. Indeed, (Gagnon & Doré, 1992) also reported that dogs sniffed more in difficult than simple object permanence tasks, and concluded that they might gather information from other sensory modalities when one is not sufficient. Studies in which dogs searched for their owners or food (Polgár et al., 2015) also found that dogs initially search by vision, but if cognitive strategies prove unsuccessful over time, they revert to search by olfaction.

In both Experiment 1 and Experiment 2, dogs slowed down, similar to the dogs in the original Watson study. However, there was one significant difference in their search behaviour. In the original Watson study, 68% of the dogs (13 out of 19) displayed direct search behaviour. In our experiments, collectively, only half of the tested dogs displayed a direct search behaviour (Collectively: 52%, Experiment 1: 51%, Experiment 2: 53% - for details see Appendix B). The remaining half searched in various indirect ways (*e.g.,* a visit to owner/experimenter, screen re-exploration, visit to locations other than screens).

Why did so many dogs search indirectly? There are two plausible explanations for the behaviour of indirect searching dogs that approached the owner after failing to find the reward behind the initial screens. Firstly, they might have approached thinking they had the reward, and they could retrieve it. Similar findings were reported by (Bräuer & Call, 2011), where dogs stayed close to the experimenter more in trials where the food reward they previously saw being placed in a container was serendipitously exchanged with another. Secondly, as mentioned above, dogs might have approached their owner to ask for help. In the human-dominated environment that dogs live in, humans (especially owners) often provide support in various domains, including access to resources such as food or toys (Marshall-Pescini et al., 2017). Hence, dogs might have expected their owner to help them and thus approached them without trying very hard to find the object by themselves. Previous studies also provide evidence that dogs socially reference humans when they are uncertain (Merola et al., 2012). It is also possible that the dogs approached the experimenter because the container she was holding smelled like the reward. However, this is unlikely, because most of the dogs inspected the container (by vision, smell, or both) before starting their search – and therefore knew that it was empty.

The higher number of indirect searching dogs in our experiments may also be attributed to differences between the original experimental procedure and the procedure of our experiments. The original study started with six *successive* visible displacement introductory trials (*i.e.,* the experimenter walks behind all screens, the reward is left behind a screen in three trials and remains in the container the other three times). In Experiment 1, we replaced the six *successive* visible displacement introductory trials with six *single* visible displacement trials (*i.e.,* the experimenter walks behind one screen, and the reward is always left behind it). It is plausible that in the original study, most dogs searched directly, simply because they adhered to the "*rules of the game*", or they mimicked the experimenter's behaviour in the introductory trials. However, if this was the case, we would anticipate an increase in the number of dogs exhibiting direct searching behaviour in Experiment 2, where we reintroduced the six successive visible displacement trials. Specifically, we expected that after seeing the experiment demonstrate a direct route, dogs would either learn that a direct search is the most effective way to search, or they would be more inclined to mimic the path of the experimenter, and thus search directly. However, despite the additional demonstration, there was no notable rise in the number of dogs engaged in direct searching. Considering this observation, we believe that the increased prevalence of indirect searching dogs in our experiments cannot be solely attributed to procedural differences compared to the original study. Further investigation is needed to uncover the exact reasons driving variations in dog search behaviour both between subjects and between experiments.

In sum, the findings of our two replication studies confirm that the Watson task is a robust method that can be used to distinguish between the use of inference and association. Importantly, since the performance of subjects is not affected by the type of reward, the task can be administered with different types of rewards, thus enabling a range of cross-species comparisons. While the compound measurement of *speed,* used in the original study accurately reflected the search behaviour of dogs, for the sake of accuracy, we propose that future studies that use this methodology would significantly benefit from a separate analysis of pace and exploration time, and automated pace analysis. Concerning the experimental procedure, given the fact that the inclusion of additional trials in Experiment 2 did not lead to more direct searching dogs, we recommend that future studies replicate the shorter procedure of Experiment 1.

1.1. Conclusion & Future Directions

While domestic dogs are renowned for their social cognition and interactions with humans, our understanding of how they perceive and reason about the physical world remains relatively limited. The two experiments reported in this thesis shed light on this unexplored area, offering two key insights. Firstly, our findings reaffirm the initial conclusion put forth by Watson et al. (2001) that dogs predominantly rely on associative strategies when searching for hidden rewards. Secondly, our experiments demonstrate that the nature of the reward, whether it be food or a toy, does not significantly impact dogs' search behaviour.

These findings carry significance from theoretical and practical standpoints and lay the groundwork for further research in canine physical cognition and comparative cognition in general. Theoretically, our findings challenge prevailing assumptions about the differential impact of food and toy rewards on canine problem-solving performance. From a practical perspective, given that the reward type doesn't seem to influence dogs' task performance, owners, trainers, and researchers can seamlessly employ food and toy rewards interchangeably to motivate dogs, whether in training contexts or experimental scenarios. Nevertheless, further investigations are warranted to explore whether dogs form distinct mental representations of these rewards or if they trigger the utilization of divergent cognitive strategies.

Our findings also contribute to the growing body of empirical evidence challenging the use of inferential reasoning by dogs during object-search tasks. While we found no conclusive evidence for inferential reasoning by exclusion, we acknowledge that this does not necessarily rule out this ability in dogs. Associative and inferential processes may coexist and may be recruited flexibly by different individuals in varying task contexts. Future research should delve deeper into these dynamics, to disentangle natural variation, individual strategy preferences, and cognitive constraints. This issue is of particular relevance to canine cognition but also holds broader implications for animal cognition research.

In addition to the aforementioned theoretical and practical implications, this thesis also makes a substantial methodological contribution to the field of comparative cognition. Our findings reaffirm the efficacy of the Watson task as a robust methodology for distinguishing between behaviour rooted in logical inference and that rooted in associative generalization. We have introduced a more rigorous analytical approach, incorporating automatic video analysis (*i.e.,* updated Watson task). The task's adaptability to various reward types renders it particularly suitable for cross-species comparisons. There is a pressing need to develop additional rigorous methodologies, and to conduct more studies with large samples, and diverse tasks, including control conditions, to rule out viable alternative non-inferential solutions.

Looking ahead, we intend to test hand-reared wolves bred and kept at the Wolf Science Center using the updated Watson Task with food as a reward. This comparative approach, contrasting the performance of pet dogs and wolves, promises valuable insights into the evolution of dog cognition in the physical domain over the course of domestication (Range & Marshall-Pescini, 2022a). While both dogs and wolves exhibit relatively poor performance in invisible displacements in object permanence tasks compared to primates, there are indications that wolves may outperform dogs in inferential skills. However, this area of research remains limited, and further exploration is warranted (for a review see (Range & Marshall-Pescini, 2022b)).

References

Amir, S., Zamansky, A., & Linden, D. van der. (2017). K9-Blyzer: Towards Video-Based Automatic Analysis of Canine Behavior. *ACM International Conference Proceeding Series*, *Part F132525*, 9:1-9:5. https://doi.org/10.1145/3152130.3152142

Anderson, D. J., & Perona, P. (2014). Toward a Science of Computational Ethology. *Neuron*, *84*(1), 18–31. https://doi.org/10.1016/J.NEURON.2014.09.005

Aust, U., Range, F., Steurer, M., & Huber, L. (2008). Inferential reasoning by exclusion in pigeons, dogs, and humans. *Animal Cognition*, *11*(4), 587–597.

Belger, J., & Bräuer, J. (2018). Metacognition in dogs: Do dogs know they could be wrong? *Learning \& Behavior*, *46*, 398–413.

Bensky, M. K., Gosling, S. D., & Sinn, D. L. (2013). The world from a dog's point of view: A review and synthesis of dog cognition research. In *Advances in the Study of Behavior* (Vol. 45). https://doi.org/10.1016/B978-0-12-407186-5.00005-7

Bleuer-Elsner, S., Zamansky, A., Fux, A., Kaplun, D., Romanov, S., Sinitca, A., Masson, S., & van der Linden, D. (2019). Computational analysis of movement patterns of dogs with ADHD-like behavior. *Animals*, *9*(12). https://doi.org/10.3390/ani9121140

Boysen, S. T. (2012). The impact of symbolic representations on chimpanzee cognition. In *Rational Animals?* https://doi.org/10.1093/acprof:oso/9780198528272.003.0022

Bräuer, J., & Call, J. (2011). The magic cup: Great apes and domestic dogs (Canis familiaris) individuate objects according to their properties. *Journal of Comparative Psychology*, *125*(3), 353–361. https://doi.org/10.1037/A0023009

Bräuer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). *Making Inferences About the Location of Hidden Food: Social Dog, Causal Ape*. https://doi.org/10.1037/0735-7036.120.1.38

Brucks, D., Soliani, M., Range, F., & Marshall-Pescini, S. (2017). Reward type and behavioural patterns predict dogs' success in a delay of gratification paradigm. *Scientific Reports*, *7*. https://doi.org/10.1038/srep42459

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition*, *9*(4). https://doi.org/10.1007/s10071-006-0037-4

Collier-Baker, E., Davis, J. M., & Suddendorf, T. (2004). Do dogs (Canis familiaris) understand invisible displacement? *Journal of Comparative Psychology*, *118*(4), 421–433. https://doi.org/10.1037/0735-7036.118.4.421

Erdohegyi, Á., Topál, J., Virányi, Z., & Miklósi, Á. (2007). Dog-logic: inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour*, *74*(4), 725–737. https://doi.org/10.1016/j.anbehav.2007.03.004

Espinosa, J. H. (2022). *The Physical World of the Dog*. University of Toronto (Canada).

Fernand, J. K., Amanieh, H., Cox, D. J., & Dorey, N. R. (2018). Any reward will do: Effects of a reverse-reward contingency on size preference with pet dogs (Canis lupus familiaris). *Learning and Behavior*, *46*(4). https://doi.org/10.3758/s13420-018-0343-0

Fiset, S., & Plourde, V. (2013). Object permanence in domestic dogs (canis lupus familiaris) and gray wolves (canis lupus). *Journal of Comparative Psychology*, *127*(2), 115–127. https://doi.org/10.1037/a0030595

Frank, H. (1980). Evolution of Canine Information Processing under Conditions of Natural and Artificial Selection. *Zeitschrift Für Tierpsychologie*, *53*(4), 389–399. https://doi.org/10.1111/j.1439-

0310.1980.tb01059.x

Fux, A., Zamansky, A., Bleuer-Elsner, S., van der Linden, D., Sinitca, A., Romanov, S., & Kaplun, D. (2021). Objective video-based assessment of adhd-like canine behavior using machine learning. *Animals*, *11*(10). https://doi.org/10.3390/ani11102806

Gagnon, S., & Doré, F. Y. (1992). Search behavior in various breeds of adult dogs (Canis familiaris): Object permanence and olfactory cues. *Journal of Comparative Psychology*, *106*(1), 58–68. https://doi.org/10.1037/0735-7036.106.1.58

Horowitz, A., & Franks, B. (2020). What smells? Gauging attention to olfaction in canine cognition research. *Animal Cognition*, *23*(1). https://doi.org/10.1007/s10071-019-01311-z

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "fast mapping." *Science*, *304*(5677). https://doi.org/10.1126/science.1097859

Karl, S., Boch, M., Zamansky, A., van der Linden, D., Wagner, I. C., Völter, C. J., Lamm, C., & Huber, L. (2020). Exploring the dog–human relationship by combining fMRI, eye-tracking and behavioural measures. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-79247-5

Krichbaum, S., & Lazarowski, L. (2022). Reward Type Affects Dogs' Performance in the Cylinder Task. *Animal Behavior and Cognition*, *9*(3). https://doi.org/10.26451/abc.09.03.03.2022

Kruijt, J. P. (1964). Ontogeny of social behaviour in burmese red junglefowl (Gallus gallus spadiceus). *Behaviour. Supplement*, *12*.

Leonardi, R. J., Vick, S. J., & Dufour, V. (2012). Waiting for more: The performance of domestic dogs (Canis familiaris) on exchange tasks. *Animal Cognition*, *15*(1). https://doi.org/10.1007/s10071-011-0437-y

Marshall-Pescini, S., Cafazzo, S., Virányi, Z., & Range, F. (2017). Integrating social ecology in explanations of wolf–dog behavioral differences. In *Current Opinion in Behavioral Sciences* (Vol. 16, pp. 80–86). Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2017.05.002

Menaker, T., Monteny, J., de Beeck, L. O., & Zamansky, A. (2022). Clustering for Automated Exploratory Pattern Discovery in Animal Behavioral Data. *Frontiers in Veterinary Science*, *9*. https://doi.org/10.3389/fvets.2022.884437

Merola, I., Prato-Previde, E., & Marshall-Pescini, S. (2012). Dogs' Social Referencing towards Owners and Strangers. *PLoS ONE*, *7*(10), e47653. https://doi.org/10.1371/journal.pone.0047653

Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Z., & Csányi, V. (2003). A simple reason for a big difference: Wolves do not look back at humans, but dogs do. *Current Biology*, *13*(9), 763–766. https://doi.org/10.1016/S0960-9822(03)00263-X

Piaget, J. (1952). *The origins of intelligence in children* (Second edition.). International Universities Press.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2). https://doi.org/10.1016/j.beproc.2010.11.007

Polgár, Z., Miklósi, Á., & Gácsi, M. (2015). Strategies used by pet dogs for solving olfaction-based problems at various distances. *PLoS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.0131610

Range, F., Hentrup, M., & Virányi, Z. (2011). Dogs are able to solve a means-end task. *Animal Cognition*, *14*(4), 575–583. https://doi.org/10.1007/s10071-011-0394-5

Range, F., & Marshall-Pescini, S. (2022a). Comparing wolves and dogs: current status and implications for human 'self-domestication.' In *Trends in Cognitive Sciences* (Vol. 26, Issue 4). https://doi.org/10.1016/j.tics.2022.01.003

Range, F., & Marshall-Pescini, S. (2022b). *Physical Cognition Part II: Learning and Reasoning*. https://doi.org/10.1007/978-3-030-98411-3_10

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6). https://doi.org/10.1109/TPAMI.2016.2577031

Riemer, S., Ellis, S. L. H., Thompson, H., & Burman, O. H. P. (2018). Reinforcer effectiveness in dogs—The influence of quantity and quality. *Applied Animal Behaviour Science*, *206*. https://doi.org/10.1016/j.applanim.2018.05.016

Sulikowski, D., & Burke, D. (2011). Movement and memory: Different cognitive strategies are used to search for resources with different natural distributions. *Behavioral Ecology and Sociobiology*, *65*(4). https://doi.org/10.1007/s00265-010-1063-4

Szetei, V., Miklósi, Á., Topál, J., & Csányi, V. (2003). When dogs seem to lose their nose: An investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaviour Science*, *83*(2). https://doi.org/10.1016/S0168-1591(03)00114-X

Triana, E., & Pasnak, R. (1981). Object permanence in cats and dogs. *Animal Learning & Behavior*, *9*(1), 135–139. https://doi.org/10.3758/BF03212035

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In *APA handbook of comparative psychology: Perception, learning, and cognition.* (pp. 643–671). American Psychological Association. https://doi.org/10.1037/0000012-029

Ward, C., & Smuts, B. B. (2007). Quantity-based judgments in the domestic dog (Canis lupus familiaris). *Animal Cognition*, *10*(1). https://doi.org/10.1007/s10071-006-0042-7

Watson, J. S., Gergely, G., Csanyi, V., Topal, J., Gacsi, M., & Sarkozi, Z. (2001). Distinguishing logic from association in the solution of an invisible displacement task by children (Homo sapiens) and dogs (Canis familiaris): Using negation of disjunction. *Journal of Comparative Psychology*, *115*(3), 219–226. https://doi.org/10.1037/0735-7036.115.3.219

Zamansky, A., Sinitca, A., Linden, D. Van Der, & Kaplun, D. (2021a). Automatic Animal Behavior Analysis: Opportunities for Combining Knowledge Representation with Machine Learning. *Procedia Computer Science*, *186*, 661–668. https://doi.org/10.1016/J.PROCS.2021.04.187

Zamansky, A., Sinitca, A., Linden, D. Van Der, & Kaplun, D. (2021b). Automatic Animal Behavior Analysis: Opportunities for Combining Knowledge Representation with Machine Learning. *Procedia Computer Science*, *186*, 661–668. https://doi.org/10.1016/J.PROCS.2021.04.187

Zamansky, A., Sinitca, A. M., Kaplun, D. I., Plazner, M., Schork, I. G., Young, R. J., & de Azevedo, C. S. (2019). Analysis of dogs' sleep patterns using convolutional neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11729 LNCS*. https://doi.org/10.1007/978-3-030-30508-6_38

Appendix A
A1 –Subject Demographics (Experiment 1)

Table 3: Experiment 1 - Overall gender and age distribution

	Overall	Food Group	Toy Group
Males (N)	27	11	16
Females (N)	29	18	11
Age - Mean	5,48	6,62	4,26
Age -SD	3,34	3,33	2,90
Total (N)	56	29	27

Table 4: Experiment 1 - Demographic details of subjects. Subject ID represents the order of testing

Subject ID	Name	Breed	FCI Group	Age	Sex
Food1	Ben	english springer spaniel	9	8	M
Food2	Noah	mix	0	1	M
Food3	Fenja	boxer	2	9	F
Food4	Waiana	boxer	2	2	F
Food5	Janosch	mix	0	4	M
Food6	Toffee	continental bulldog	2	5	F
Food7	Sixtus	petit brabançon	9	4	M
Food8	Diva	english bulldog	2	8	F
Food9	Suri	mix	0	5	F
Food10	Kyra	dacshund mix	4	4	F
Food11	Suri	mix	0	5	F
Food12	Eowyn	terrier mix	3	3	M
Food13	Leia	beagle	6	9	F
Food14	Nikki	Bernese mountain dog	2	4	F
Food15	Junie	border collie	1	5	F
Food16	Gina	German shepherd	1	1	F
Food17	Hugo	Bernese mountain dog	2	11	M
Food18	Aramis	golden retriever	8	7	M
Food19	Darius	Bedlington Terrier	3	12	M
Food20	Jessie	Mix	0	11	F
Food21	Kiki2	Mix	0	6	F
Food22	Holly6	Beagle	6	5	F
Food23	Izy	Podenco	5	9	F
Food24	Jack6	Mix	0	14	M
Food25	Honey4	Labrador	8	12	F
Food26	Gismo4	Mix	0	6	M
Food27	Sheila	Border Collie	1	9	F
Food28	Sunny11	Mix	0	8	M
Food29	Pamina2	Mix	0	5	F

Subject ID	Name	Breed	FCI Group	Age	Sex
Toy1	Samira	Yorki	3	6	F
Toy2	Niki	jack russell	3	4	M
Toy3	Finley	shetland sheepdog	1	3	M
Toy4	Shadow	border collie	1	6	M
Toy5	Onyx	German Shepherd	1	2	M
Toy6	Ronja	Riesenschnauzer	2	5	F
Toy7	Arkani	Belgian shepherd dog	1	6	M
Toy8	Emilio	golden retriever	8	5	M
Toy10	Ookami	White swiss shepherd dog	8	4	M
Toy11	Befana	mix	1	6	F
Toy12	Fossy	mix	0	3	M
Toy13	Fynn	mix	0	1	M
Toy14	Destiny	coonhound	6	8	F
Toy15	Pandora	Am. staffordshire terrier	3	3	F
Toy16	Django2	Am. staffordshire terrier	3	5	M
Toy17	Emma17	Shetland Sheepdog	1	1	F
Toy18	Roxy7	Am. Staffordshire Terrier	3	2	F
Toy19	Jolly4	White Springer Spaniel	8	11	F
Toy20	Baghira	Australian Shepherd	1	10	M
Toy21	Charlie	Australian Shepherd	1	7	M
Toy22	Snow	Australian Shepherd	1	1	M
Toy23	Wilma2	Mix	0	1	F
Toy24	Cameron	Border Collie	1	9	M
Toy25	Bubbles	Mix	0	1	F
Toy26	Gryffindor	Border Collie	1	3	M
Toy27	Baloo	Australian Shepherd	1	1	M
Toy28	Flo	Mix	0	1	F

A2 – Subject Demographics (Experiment 2)

Table 5: Experiment 2 - Overall gender and age distribution

	Overall	Food Group	Toy Group
Males (N)	35	13	22
Females (N)	37	26	11
Age - Mean	5,86	6,79	4,76
Age -SD	3,50	3,39	3,30
Total (N)	72	39	33

Table 6: Experiment 2 - Demographic details of subjects. Subject ID represents the order of testing

Subject ID	Name	Breed	FCI Group	Age	Sex
Food1	Emma10	Labrador retriever	8	5	F
Food2	Defi	Mix	0	3	F
Food3	Joe	Irish Terrier	3	2	M
Food4	Aofie	Irish Glen of Imaal Terrier	3	8	M
Food5	Daisy3	Mix	0	6	F
Food6	Dunni	Miniature Pinscher	2	1	F
Food7	Coco3	Siberian Husky	5	6	F
Food8	Flamme	Pyrenean Sheepdog	1	13	M
Food9	Jolie	Mix	0	11	F
Food10	Lillybet	Samojede	5	6	F
Food11	Ignaz	Fox Terrier (Smooth)	3	1	M
Food12	Leo2	Mix	0	7	M
Food13	Lea3	Parson Russel Terrier	3	7	F
Food14	Duke4	Other	0	6	M
Food15	Eleanor	Mix	0	12	F
Food16	Lucy14	Beagle	6	3	F
Food17	Toulouse	Mix	0	2	M
Food18	Yussi	Mix	0	11	F
Food19	Denzel	Am. Staffordshire Terrier	3	12	M
Food20	Sugar2	Am. Staffordshire Terrier	3	13	F
Food21	Jessy8	Shetland Sheepdog	1	8	F
Food22	Stemo	Mix	1	6	M
Food23	Loki4	Mix	1	7	M
Food24	Pepper	Whippet Mix	0	8	F
Food25	Saphir	White Swiss Shepherd	1	5	F
Food26	Joline	Whippet	10	4	F
Food27	Emily8	Mix	0	5	F
Food28	Josephine	Mix	0	4	F
Food29	Nils	Mix	0	9	M

Subject ID	Name	Breed	FCI Group	Age	Sex
Food30	Cappuccina	Border Collie	1	6	F
Food31	Luzi	Spanish Greyhound	10	7	F
Food32	Stani	Airedale Terrier	3	6	M
Food33	Toffee3	Continental Bulldog	2	7	F
Food34	Shelby	English Bulldog	2	5	F
Food35	Majomka	Mix	0	12	F
Food36	Kaja	Mix	0	11	F
Food37	Waleria	German Pinscher	2	1	F
Food38	Kobe	Mix	0	8	M
Food39	Mocca2	Labrador Retriever	8	11	F
Toy1	Finne-Failbhe	Irish Glen of Imaal Terrier	3	1	F
Toy2	Manni2	Pyrenäen Sheep dog	1	5	M
Toy3	Jimmy5	Mix (Terrier)	0	6	M
Toy4	Linus3	Border Collie	1	3	M
Toy5	Loki2	Australian Shepherd	1	4	M
Toy6	Tango2	Australian Shepherd	1	1	M
Toy7	Katica	Hungarian Greyhound	10	11	F
Toy8	Lennox5	Mix	0	2	M
Toy9	Zserbo	Kleiner Münsterländer	7	1	M
Toy10	Unico	Shetland Sheepdog	1	7	M
Toy11	Wilson	Golden Retriever	8	2	M
Toy12	Jasper	Labradoodle	0	11	M
Toy13	Riu	Australian Shepherd	1	3	F
Toy14	Joy10	Parson Russel Terrier	3	6	F
Toy15	Yuna	Flat Coated Retriever	8	6	F
Toy16	Tenya	Collie Rough	1	5	F
Toy17	Lenny	Border Collie	1	12	M
Toy18	Mogli5	Border Collie	1	1	M
Toy19	Nuala	Mix	0	7	F
Toy20	Leonard	Mix	0	6	M
Toy21	Alou	Australian Shepherd	1	5	F
Toy22	Willo	Flatcoated Retriever	8	9	M
Toy23	Roxy5	Irish Red and White Setter	7	6	F
Toy24	Maylo	Mix	0	5	M
Toy25	Gandalf	Andalusian Hound	5	2	M
Toy26	Ultimo	Border Collie	1	12	M
Toy27	Quinn	Weimaraner	7	1	M
Toy28	Crash	Australian Shepherd	1	3	M
Toy29	Henry5	Golden Retriever	8	6	M
Toy30	Oonagh	Labrador	8	1	F
Toy31	Minou	Mix	0	4	F
Toy32	Punch	Mix	0	2	M
Toy33	Roku	Mudi	1	1	M

Appendix B
Descriptive Statistics

Tables reporting sample size (per group), total number of subjects that were successful in each set of trials of the procedure, number and percentage of subjects that searched in a direct/indirect way in the test trial (per group and total). The final sample size for the statistical analysis is the one reported under the Test Trial column.

Table 7: Descriptive Statistics – Experiment 1

		Successful Subjects (N)				**Direct Search**	**Indirect Search**
Group	**Sample Size**	**Screening Trials**	**Visible Displacement Trials**	**Invisible Displacement Trials**[2]	**Test Trial**	**N (% within group)**	**N (% within group)**
Food	29	29	22	-	22	11 (50%)	11 (50%)
Toy	27	27	21	-	21	11 (52%)	10 (48%)
Total	**56**	**56**	**43**	**-**	**43**	**22 (51%)**	**21 (49%)**

Table 8: Descriptive Statistics – Experiment 2

		Successful Subjects (N)				**Direct Search**	**Indirect Search**
Group	**Sample Size**	**Screening Trials**	**Visible Displacement Trials**	**Invisible Displacement Trials**	**Test Trial**	**N (% within group)**	**N (% within group)**
Food	39	36	27	27	24	13 (54%)	11 (46%)
Toy	33	33	28	28	21	12 (57%)	9 (43%)
Total	**72**	**69**	**55**	**55**	**45**	**25 (56%)**	**20 (44%)**

---

[2] Invisible displacement trials were only run in Experiment 2

Appendix C

C1 – Behavioural Ethogram

Table 9: Behavioural ethogram describing six potential types of screen exploration and related behaviours (re-exploration, pass, lookout, turnaround).

Code	Type of Behaviour	Description
Exploration-Behaviour1	Olfactory exploration of the area behind a screen –with physical contact (*e.g.,* sniffing)	-Nose inside the exploration area behind the screen -Nose in physical contact with the floor or the screen or sniffing floor/screen -Any head orientation
Exploration-Behaviour2	Visual exploration of the area behind a screen – without physical contact (*e.g.,* glancing)	-Nose inside the exploration area behind the screen -No physical contact between the nose and the floor or the screen -Head orientation towards the interior of the screen
Exploration-Behaviour3	Olfactory or visual exploration of the area in front of a screen - with physical contact	-Nose touching front of the screen -Nose in physical contact with the floor or the screen or sniffing floor/screen -Head orientation towards the front of the screen
Exploration-Behaviour4	Visual exploration of the area behind a screen by climbing up from the front	-Dog climbing up on the screen from the front -Nose inside the exploration area behind the screen -Head turned down, towards the interior of the screen
Exploration-Behaviour5	Visual exploration of the area behind a screen, while being in the wider area behind that screen– without physical contact with the floor (only applicable to Screens 1 and 3)	-Nose inside the wider exploration area behind the screen -Head orientation towards the interior of the screen -No physical contact between the nose and the floor
Exploration-Behaviour6	Olfactory exploration of the area behind a screen, while being in the wider area behind that screen – with physical contact with the floor (only applicable to Screens 1 and 3)	-Nose inside the wider exploration area behind the screen -Physical contact between the nose and the floor -Head orientation towards the interior of the screen
Re-Exploration	Re-exploration of a screen that has already been explored	-Exploration of a screen that has already been explored, prior to finishing exploration of all screens

Code	Type of Behaviour	Description
Pass	Travelling in front or behind, or coming near to, a screen without exploring it	-Dog is in the proximity of a screen, none of the exploration behaviours (Exp-Beh 1-6) is fulfilled, and one of the 3 following behaviours is fulfilled: -Coming within 20cm of the front of a screen with the head oriented towards the screen but not meeting any exploration behaviour criteria (due to no physical contact with screen/floor); dog then travels away from this screen to a different screen -Walking behind a screen from one side to the other side (*i.e.,* nose crossing both the left and right sides of the triangle demarcating the exploration area of that screen), but not meeting any of the screen exploration behaviour criteria -Traveling towards a screen and looking in the direction of the screen interior (*e.g.,* glancing towards it) but from too far away to meet exploration behaviour criteria (due to the nose not being inside the relevant exploration area when looking); the dog subsequently travels away from this screen to a different screen.
Lookout	Short (<2 sec) gap during exploration (*e.g.,* briefly looking away from the screen interior)	-Dog is exploring a screen (according to Exp-Beh 1-6) -Dog briefly does not meet criteria for any of Exp-Beh 1-6 (for less than 2 seconds) and then meets criteria again, *i.e.,* starts exploring again
Turnaround	Turning around behind a screen to change the direction of travel and subsequently travel to another screen	-Dog ends exploration of screen (no longer meets Exp-Beh 1-6) -Dog then turns around behind the screen in order to subsequently travel to another screen in a different direction; while doing so, they briefly meet Exp-Beh2 (head briefly oriented towards the interior of the screen) - this is not counted as further exploration of the screen, since meeting Exp-Beh2 here is part of travel to the next screen

C2- Exploration Time Coding

Three exploration time values are coded for each subject (E1, E2, E3).

Start of exploration: Exploration time starts once one of the exploration criteria (see Ethogram) is fulfilled. During exploration, a dog might perform a behaviour that does not meet exploration behaviour criteria (*i.e.,* look away from the screen interior). Depending on the duration of this behaviour – referred to as a *gap*, we either code a *lookout* or *the end of the exploration.*

End of exploration: If none of the exploration criteria are fulfilled for a period of 2 seconds or longer, then we code the *end* of that exploration. Exploration time ends before a gap, and at the same time, we code the *start* of travel time.

Special cases

Lookout: A gap in the exploration of a screen that is less than 2 seconds (*e.g.,* briefly glancing away from the screen interior). The duration of a lookout is included in the overall exploration time *i.e.,* the recording of exploration time continues from the start of exploration behaviour before the lookout until the end of exploration behaviour following the lookout; the duration of the lookout is not subtracted from the overall exploration time. Gaps longer than 2 seconds mark the end of exploration time for the given screen.

Re-exploration: If a dog re-explores a screen that was already explored, then this is coded as a *re-exploration*. The travel time calculation for the next unexplored screen includes the time of intermediate re-exploration trips but does not include the time dedicated to re-exploring the screens. Likewise, travel distance calculation includes the distance covered in the intermediate re-exploration trips, but not the distance covered while re-exploring them. This ensures that re-explorations are reflected in the final pace calculations.

Subsequent re-exploration of the same screen: If a dog finishes exploring a screen, and then begins exploring the same screen again, after a gap that is 2 seconds or longer, then this is coded as a *re-exploration of the same screen.* Exploration time for that screen is calculated by adding together the time of the original exploration and the subsequent re-exploration(s). The *gaps* between these explorations are coded as travels – but these travel time scores are not included in the three final travel time scores.

C3 - Travel Time and Travel Distance Coding

Three travel time values (T1, T2, T3) and three travel distance values (D1, D2, D3) are coded for each subject.

Start of travel time: For travels between the starting point and a screen, travel time begins as soon as the dog starts to move, after receiving the search command. For travels between screens (screen A – screen B), travel time begins once the exploration of the first screen (screen A) ends (*i.e.,* no exploration criterion is fulfilled for more than 2 sec).

End of travel time: Travel time ends once the dog starts exploring a screen (*i.e.,* one of the exploration criteria is fulfilled for more than 2 sec).

Special cases

Turnaround: If a dog finishes exploring a screen, and then carries out a turn behind that screen to subsequently travel to another screen, we code a *turnaround.* Exploration time ends before the beginning of the turning when none of the exploration behaviours are being fulfilled any longer. This is true even if the process of turning around involves the dog's head being (briefly) oriented towards the interior of the screen (and therefore one of the exploration criteria is met very briefly) while turning. A *turnaround* is considered part of the post-exploration travel time score, rather than part of the screen exploration time score. Therefore, travel time starts when exploration time ends (before the turnaround starts). The duration of the turnaround is therefore included within the duration of the travel to the next screen.

Pass: While traveling between two screens, a dog might come near another screen without exploring it (*e.g.,* the dog is in front or behind a screen, but none of the behaviours Exp-Beh 1-6 is fulfilled). This behaviour does not influence travel time or travel distance calculations – but it is used to determine if a travel trajectory is direct or indirect.

C4 - Pace Coding

Pace is measured in units of time (measured in seconds) per unit of distance (measured in pixels) and is calculated as follows:

$$\text{Pace} = \frac{\sum\,(\mathrm{T1,T2,T2})}{\sum\,(\mathrm{D1,D2,D3})}$$

Special cases

Indirect search: For dogs that follow an *indirect* trajectory involving re-explorations, total travel time is calculated by summing up all intermediate travel times. Likewise, total travel distance is calculated by summing up all intermediate travel distances. For example, if a dog travels to screen 1, explores it, then travels to screen 2, explores it, and returns to screen 1 before travelling to screen 3 to explore it, the total travel time will be the sum of the following travel times: starting point to screen 1, screen 1 – screen 2, screen 2-screen 1, screen 1- screen 3. Likewise, total travel distance will be the sum of the corresponding distances.

Appendix D

D1 - Linear Mixed Models (Experiment 1)

Model 1

RESPONSE:	Log (Pace)
DISTRIBUTION	Gaussian
KEY PREDICTORS:	Reward * ScreenNumber
CONTROL PREDICTORS	Experimenter
RANDOM INTERCEPT EFFECTS	subject
RANDOM SLOPES?	yes
CORRELATION BETWEEN R.E.	yes
NR. OBSERVATIONS	129
FULL MODEL:	log.eff.pace.auto ~ Reward * z.ScreenNumber + experimenter + (1 + z.ScreenNumber | subjectID)
NULL MODEL:	log.eff.pace.auto ~ experimenter + (1 + z.ScreenNumber | subjectID)
Full-null model comparison **LIKELIHOOD RATIO TEST:**	
X^2	7,999
DF	3
***P*-VALUE**	0,046
Variation explained:	
$R^2_{MARGINAL}$ **(FIXED EFFECTS PROP. VAR. EXPL.)**	0,276
$R^2_{CONDITIONAL}$ **(FIXED+RANDOM EFFECTS PROP. VAR. EXPL.)**	0,706

Model 2

RESPONSE:	Log (Explorationtime)
DISTRIBUTION	Gaussian
KEY PREDICTORS:	Reward * ScreenNumber
CONTROL PREDICTORS	experimenter
RANDOM INTERCEPT EFFECTS	subject
RANDOM SLOPES?	yes
CORRELATION BETWEEN R.E.	no
NR. OBSERVATIONS	129
FULL MODEL:	log.exp.tim ~ Reward * z.ScreenNumber + experimenter + (1 + z.ScreenNumber || subjectID)
NULL MODEL:	log.exp.tim ~ experimenter + (1 + z.ScreenNumber || subjectID)
Full-null model comparison LIKELIHOOD RATIO TEST:	
X^2	5,748
DF	2
***P*-VALUE**	0,125
Variation explained:	
R^2_{MARGINAL} **(FIXED EFFECTS PROP. VAR. EXPL.)**	0,126
$R^2_{\text{CONDITIONAL}}$ **(FIXED+RANDOM EFFECTS PROP. VAR. EXPL.)**	0,148

D2 - Linear Mixed Models (Experiment 2)

Model 1

RESPONSE:	Log (Pace)
DISTRIBUTION	Gaussian
KEY PREDICTORS:	Reward* ScreenNumber
CONTROL PREDICTORS	Search behaviour + Experimenter
RANDOM INTERCEPT EFFECTS	subject
RANDOM SLOPES?	yes
CORRELATION BETWEEN R.E.	no
NR. OBSERVATIONS	135
FULL MODEL:	log.eff.pace.auto ~ Reward* z.ScreenNumber + searchBehaviour + experimenter + (1 + z.ScreenNumber || subjectID)
NULL MODEL:	log.eff.pace.auto ~ searchBehaviour + experimenter + (1 + z.ScreenNumber || subjectID)

Full-null model comparison **LIKELIHOOD RATIO TEST:**	
X^2	22,714
DF	3
***P*-VALUE**	0,000
Variation explained:	
R^2_{MARGINAL} **(FIXED EFFECTS PROP. VAR. EXPL.)**	0,145
$R^2_{\text{CONDITIONAL}}$ **(FIXED+RANDOM EFFECTS PROP. VAR. EXPL.)**	0,309

Model 2

RESPONSE:	Log (exploration time)
DISTRIBUTION	Gaussian
KEY PREDICTORS:	Reward* ScreenNumber
CONTROL PREDICTORS	Search behaviour + Experimenter
RANDOM INTERCEPT EFFECTS	subject
RANDOM SLOPES?	yes
CORRELATION BETWEEN R.E.	no
NR. OBSERVATIONS	135
FULL MODEL:	log.exp.tim ~ Reward* z.ScreenNumber + searchBehaviour + experimenter + (1 + z.ScreenNumber || subjectID)
NULL MODEL:	log.exp.tim ~ searchBehaviour + experimenter + (1 + z.ScreenNumber || subjectID)

Full-null model comparison **LIKELIHOOD RATIO TEST:**	
X^2	4,470
DF	3
***P*-VALUE**	0,215
Variation explained:	
R^2_{MARGINAL} **(FIXED EFFECTS PROP. VAR. EXPL.)**	0,083
$R^2_{\text{CONDITIONAL}}$ **(FIXED+RANDOM EFFECTS PROP. VAR. EXPL.)**	0,331