Fall 1-6-2022

An Exploration of the Contextual Interference Effect on Trained Trick Retention in Companion Dogs (Canis lupus familiaris)

Madeline Messina
CUNY Hunter College

An Exploration of the Contextual Interference Effect on Trained Trick Retention in Companion Dogs (*Canis lupus familiaris*)

Maddie Messina

Submitted in partial fulfillment

of the requirements for the degree of

Master of Arts in

Animal Behavior and Conservation, Hunter College

The City University of New York

Fall 2021

01/06/2022	Dr. Sarah-Elizabeth Byosiere
Date	Thesis Sponsor
01/06/2022	Dr. Gal Ziv
Date	Second Reader

Abstract

The ability to enact behavior change is pivotal to dog training success. Currently, there are not many studies informing the best practices for doing so. This particularly impacts the working and companion dog industries, as there is support for the effect of positive behavior change on both human-dog relationships and training program success rates. This thesis sought to enhance current training practices by applying a human motor skill learning theory, the contextual interference effect (CI), to a trick-training paradigm with companion dogs. To test this theory in dogs, we trained 17 dogs to perform three skills of varying complexity in a blocked format (low CI) and a random format (high CI). One day later, dogs returned for a retention test. Throughout the acquisition and retention phases, several measures were used to assess dogs' learning. Disengagement from the handler was used to assess dogs' cognitive load. In humans, research shows that practicing skills in random order, as compared to practicing skills in blocked order, improves learning of those skills. Despite this, we found no significant differences between dogs who practiced three tricks in random or in blocked order during training and retention testing ($p > .05$). The type of training dogs received also did not significantly impact the dogs' disengagement scores ($p > .05$). This study is the first to apply the CI effect to dog trick training. While no evidence of the CI effect was found, this thesis provides a framework for future studies interested in examining the effect of CI on dog training schedules. Further examination into the CI effect on dog training could have implications for increased retention of trained skills. Similarly, further research investigating disengagement as a tool for measuring cognitive load could provide insight into the efficacy of dog training techniques.

Keywords: dogs, contextual interference, learning, training, companion dogs, working dogs

Acknowledgements

I would like to extend heartfelt gratitude to my advisors Dr. Sarah-Elizabeth Byosiere and Dr. Gal Ziv. This thesis would not have been possible without their unending support and guidance. I am especially thankful for their patience with me and the pauses I had to take throughout this project as life got in the way. Thank you to the other members of the Hunter faculty who also supported this thesis: Dr. James Gordon, Dr. Martin Chodorow, Dr. Diana Reiss, and Julie Hecht. I would be remiss without acknowledging the time and effort committed to the project by my fabulous coders and fellow Hunter Students, Louis Lin and Samantha Smith. Thank you both for being patient and thoughtful as we navigated the behavior coding aspect of this project together.

Behind the scenes, my family and friends have been an un-ending source of support and for that I am forever thankful. A special thank you to my parents, for continuing to push me to exceed my potential and for their support of my love of animals, even if at times the depth of that love is hard for them to fathom. Thank you also to my boyfriend, Steve, who has been my sounding board and emotional support person throughout the entire process.

Lastly thank you to all the guardian and dog teams who took the time out of their lives to participate in this project. I know that it was not always easy for them to make the trek to the Thinking Dog Center, but they did it anyways. I am indebted to the humans and dogs alike for their interest in this thesis and their commitment to furthering canine science.

Introduction

Positioning Dogs as an Economic Investment

Domestic dogs (*Canis lupus familiaris*) possess unique cognitive adaptations for the wide variety of roles they play in society, each role posing its own challenges (Arden et al., 2016). Most modern-day canines, who live and work with humans on a regular basis, can be categorized as belonging to one of two roles: companion or working animals. Of the two, working dogs probably face the most intense challenges. They guard other animals, herd, help blind and physically frail humans, aid in search and rescue operations, and are now being used to identify diseases in humans (Brady et al., 2018; Maclean & Hare, 2018). For some of these roles, genetic predisposition is of utmost importance. For other roles, there is a good deal of effort involved in preparing these animals for their future careers (Bogaerts et al., 2019; Bray et al., 2019; Bray et al., 2021).

Working dogs are often selectively bred for temperament and functionality in fulfillment of specific roles. However, breeding alone cannot guarantee their successful placement. In fact, the stark number of rejections (estimated at around 50-70% in assistance dogs) common to working dogs in breeding and training programs has spurred the implementation of a variety of temperament and cognitive battery tests to predict their future success more accurately (Brady et al., 2018; Bray et al., 2021, Cobb et al., 2015). Not only do these dogs undergo a battery of exams, but they also spend two or more years in training if you include the valuable time spent with foster and puppy raising families when developing and maturing (Bogaerts et al., 2019; Bray et al., 2019; Bray et al., 2021). Working dogs can be a hefty investment. Costing approximately $50,000 to train a single dog (and handler), it’s no wonder programs involved with breeding and training working dogs are eager to increase their return of investment (Cobb et al., 2015; Otto et al., 2021).

Alongside working dogs, companion dogs also play an important role in society. These dogs can provide comfort and health benefits to their guardians (Utz, 2014). However, obedience issues, house soiling, destructiveness, and various anxiety and fear-related behaviors can detract from the benefits of dog ownership (Bennett et al., 2007). Because of these behavioral issues, many guardians seek out the help of a professional dog trainer in dealing with dog behavior woes. The pet training services industry is currently valued at around 1 billion dollars and estimated to grow to 1.5 billion dollars by 2026 (Research and Markets, 2021). Most guardians would agree that owning a well-behaved pet can enrich their lives (Bennett et al., 2007). In fact, positive behavior change in companion dogs is correlated with satisfaction in their human caretakers (Bennett et al., 2007).

As demonstrated above there is a clear link between behavior change and success in the working dog and companion dog industries. In working dogs, behavior change is paramount to their successful graduation from a training program (Bray et al., 2021; Hall et al., 2021). In companion dogs, guardians’ satisfaction with their dogs is correlated with behavior change and dog relinquishment is correlated with behavior problems, regardless of their enrollment in training classes (Bennett et al., 2007; Kwan & Bain, 2013). This evidence suggests that training outcome affects the future success of dogs in their companion and working dog roles. Given the link between behavior change and success in the working dog and companion dog industries, behavior and learning research should be prioritized to increase training program return of investment.

The Dog Training Field: What is Already Known

Given the time and monetary investment required from most dog training facilities, it is important to understand the effects of different training methods on learning. What evidence is informing the disciplines taught and guaranteeing return of investment for guardians and businesses? Dog training research thus far can be broken down into three main fields: (1) efficacy of training method considers the use of operant conditioning, social learning, and reward- vs. punishment-based methods in training success (China et al., 2020; Deldalle & Gaunet, 2014; Fugazza & Miklósi, 2014, 2015; Schilder & van der Borg, 2004; Stellato et al., 2019; Ziv, 2017); (2) efficacy of marker usage examines operant conditioning paradigms and the role that the marker and a food reward play in training success (Dorey & Cox, 2018; Feng et al., 2016, 2018; Gilchrist et al., 2021); and (3) the effect of training schedule on learning investigates how the rate and duration of training sessions plays a role in training success (Demant et al., 2011; Meyer & Ladewig, 2008).

Of the three, the effect of training schedule on learning seems to be the least informed field with only two notable studies that have emerged in the past 20 years. Meyer and Ladewig (2008) compared the effects of daily (5 days a week) or weekly (once per week) training sessions on the acquisition of a paw-targeting task in 18 laboratory Beagles ranging from 1-3 years old. Dogs were trained daily/weekly until they acquired the most advanced level of the paw-targeting task. The task was taught via operant conditioning and broken down into four steps. Each session consisted of 15 training trials. Dogs who were trained once weekly ($M = 6.66$, $SD = 1.50$) acquired the final training step in fewer sessions than dogs trained daily ($M = 9.00$, $SD = 2.06$).

In a similar study, Demant et al. (2011) examined the effect of once-twice weekly training sessions versus daily sessions on the acquisition of an obedience task, segmented into 12 levels, in 44 laboratory Beagles ranging from 5 months to 6 years old. The researchers also studied the effect of training session duration on task acquisition. Dogs were assigned to four groups: W1 (1-2 trainings per week/1 session per training), W3 (1-2 trainings per week/3 sessions per training), D1 (daily trainings/1 session per training), and D3 (daily trainings/3 sessions per training). The dogs in this study practiced for 18 sessions to acquire the obedience skill regardless of grouping. Each session consisted of six trials. Retention of the task was tested 4 weeks after the 18th session. Dogs who were trained once or twice per week and only received one session of training per day improved faster than dogs who received daily training sessions and dogs who received three training sessions in a row, either weekly or daily. The level of the obedience skill that was reached during acquisition predicted the obedience skill level demonstrated in the retention test.

The two abovementioned studies are unique because, to our knowledge, there are no other studies suggesting how often or how long dogs should train for. However, the caveat of relying on the information provided by the Demant et al. (2011) and Meyer & Ladewig (2008) studies is they lack real-life applicability. In both studies, dogs were trained on only one skill. In reality, working and companion dogs alike often undergo training for more than one skill at a time. Some questions that arise as a result are: Should one skill be trained per session? If not, how will training multiple skills in one session affect learning? Does training multiple skills per session affect the recommended rate and length of training sessions per week? Due to the lack of research informing this field, it is not unreasonable to examine the literature in human learning to identify techniques that could advance the acquisition of trained tasks in dogs. This is especially true given that within the past few years, researchers have placed increasing emphasis on human cognition in direct comparison to domestic dogs (Lea & Osthaus, 2018; Maclean et al., 2017).

More specifically, identifying the shared aspects of dogs' and humans' socio-cognitive abilities (Lea & Osthaus, 2018; Maclean et al., 2017).

What is Contextual Interference

While there may not be a direct comparison between the rate and length of multi-skill training sessions between dog and human learning, there is evidence from the human literature that could be used to inform the organization of individual training sessions in dogs. Since the work of Battig (1972) in verbal learning, what is now known as contextual interference has become an important topic in motor skill learning research. In their seminal work, Shea and Morgan (1979) reported the effects of contextual interference on a motor skill task. Contextual interference (CI) refers to the interference resulting from the execution of several skills at a time in a practice context (Pauwels et al., 2014). This interference leads to increased cognitive effort, which in the case of CI, is defined by the level of mental processing required by a performer to prepare for a task (Farrow & Buszard, 2017). In humans, when manipulating CI levels, researchers typically sort study participants into a 'blocked' group and a 'random' group (Shea & Morgan, 1979). The random group is required to switch between several similar skills within a specific time period or number of trials. The blocked group, on the other hand, will typically focus on one of several skills at a time. They will practice each skill for several minutes or trials in a row before moving on to the next skill (Shea & Morgan, 1979). Practicing tasks in blocked order (i.e., performing several repetitions of the same task before moving to another) is associated with lower CI, whereas practicing tasks in random order (i.e., performing random repetitions of all tasks to be learned not in a specific order) is associated with higher CI. A large body of research suggests that high CI, resulting from practicing variations of a similar motor task in random order, leads to reduced performance during practice but to better performance in retention and transfer tests (Ollis et al., 2005; Pauwels et al., 2014). Comparatively low CI, associated with practicing one skill at a time in a blocked manner, results in higher performance in practice but lower performance in retention and transfer testing.

In CI paradigms, retention and transfer tests are typically conducted between one to a few days after training is complete (Farrow & Buszard, 2017). The retention test usually requires participants to perform the skills exactly as they were trained in acquisition over a smaller number of trials. The transfer test typically requires participants to perform some variation on the skills they acquired during training or requires the participants to perform the same skills as the retention test in a novel condition. In an example of a CI paradigm involving the acquisition of badminton serves (Goode & Magill, 1986), participants were required to practice three different serves, over 36 trials, from the right-hand side of the court during acquisition. During retention testing one day later, they were required to perform the same three serves over 18 trials. The transfer test took place immediately after the retention test, in which participants performed the three different serves over 18 trials but serving from the left side of the court instead.

There are currently three contested theories explaining the mechanisms behind the CI effect. The first is the forgetting-reconstruction hypothesis, which is based on the idea that high CI resulting from random practice causes the person to forget the specifics of each task and reconstruct the task's action plan every time they switch between tasks (Lee & Magill, 1985). The second is the elaboration hypothesis, which suggests that persons training in a random format are required to elaborate further on each task in their mind to have a more distinct understanding of each task being performed (Shea & Morgan, 1979). The third is the implicit learning hypothesis, which proposes that increased retention performance due to random practice results from unconscious learning processes and requires less working-memory capacity

(Rendell et al., 2010). Of the three, there does not seem to be one prevailing theory, leaving the field of CI wide open for further investigation and application of differing research designs (Farrow & Buszard, 2017).

Specifics of Contextual Interference Studies

Studies on CI are typically conducted in laboratory or in applied settings. Most laboratory research can be narrowed down to verbal learning or fine motor skill practice (Battig, 1972; Lin et al., 2018; Pauwels et al., 2014; Shea & Morgan, 1979). Some researchers argue the results of laboratory work have implications for sports practice (Pauwels et al., 2014; Shea & Morgan, 1979). However, in a typical performance setting, sports players must complete several different skills at any given time, which often require the use of multiple motor patterns and whole-body movements (Barreiros et al., 2007; Wulf & Shea, 2002). Given this, several reviews purport that evidence based on the study of small variations to singular motor patterns cannot be applied to sports practice, thus, the call for more studies in applied settings (Barreiros et al., 2007; Farrow & Buszard, 2017; Wulf & Shea, 2002).

A pattern is emerging in recent reviews of CI in applied settings which may predict the likelihood of finding a CI effect on outcome variables. Barreiros et al. (2007) found studies investigating propulsive actions, defined by throwing, kicking, or batting sports tasks, to be less likely to provide evidence for the effect of CI on learning. In an exemplary study of golfing (Brady, 1997), participants were randomly assigned to blocked or random groups and were asked to practice four separate golf swings (drive, middle distance iron, pitch, and chip shot). The blocked group practiced each swing 15 consecutive times before moving on to the next; the random group practiced each swing 15 times randomized across 60 trials. One week later, retention of the practice was tested by recording the number of shots required to reach the green during an 18-hole round of golf. No significant differences were found between the blocked and random groups.

On the other hand, studies examining non-propulsive actions such as snowboarding skills, did show positive CI effects in retention tests (Barreiros et al., 2007). For example, Smith (2002) allocated twenty participants into blocked and random groups and investigated their ability to perform 180 degree turns on snowboards. The two turns being practiced were toe-to-heel turns and heel-to-toe turns. Blocked participants practiced four toe-to-heel turns in a row before moving on to four heel-to-toe turns, while the random group alternated between turns during acquisition. One week later, participants were scored on four trials of each turn. All participants were tested in the blocked manner regardless of acquisition grouping. Random practice resulted in higher scoring turns during retention than blocked practice. In several reviews, skill complexity is considered one of the potential factors for the differences seen between studies examining propulsive and non-propulsive actions (Barreiros et al., 2007; Farrow & Buszard, 2017; Wulf & Shea, 2002).

Within both laboratory and applied studies, skill complexity is an important factor of interest (Farrow & Buszard, 2017). As skill complexity increases so does cognitive effort. CI is associated with increased cognitive effort which is in turn correlated with increased learning. As mentioned previously, cognitive effort or the amount of cognitive effort required which is cognitive load, is defined by the level of internal processing the performer does to prepare for a task. At a certain level of task difficulty, CI no longer has a positive effect on learning and becomes less advantageous for the learner. This is likely due to the combined cognitive load of both CI and skill complexity on the learner (Wulf & Shea, 2002). Task difficulty can be defined nominally and functionally. Nominal task difficulty refers to the elements of the task that make it difficult unrelated to the environment the task is being performed in or to the skill level of the person performing that task (Guadagnoli & Lee, 2004). Functional task difficulty ascribes the difficulty of a task in relation to the person performing it (e.g., their individual skill level) and to the environment said task is performed in (Guadagnoli & Lee, 2004). To assess the level of task difficulty at which CI is no longer beneficial, nominal and functional task difficulty must be controlled and varied in a deliberate manner. This way it is clear how change in one affects the other. Deliberate consideration of both nominal and functional task difficulty could then play a role in the success of finding the CI effect in applied designs (Barreiros et al., 2007).

Our Aim

The purpose of the current study was to investigate CI in dog training with respect to task complexity. As previously mentioned, the need for research guiding training methods in pet and working dogs is clear. This thesis attempted to examine the learning process in dogs in a manner that was applicable to a real-life setting. The dogs in this study were taught and tested at the Thinking Dog Center. At the time, the Center welcomed dogs and guardians on a weekly basis. This meant that the Center for most dogs was a novel environment, not unlike any other training facility or environment in which training might be conducted. In this way, our design was as ecologically relevant as possible. We also aimed to investigate the effect of nominal task complexity on skill acquisition in the presence of CI by incorporating three skills, increasing in difficulty, during acquisition and retention. These skills (listed in order of difficulty – easiest first) were called “chin rest”, in which the dog had to rest their chin on the trainer’s hand, “platform”, the dog had to step onto a platform, and “spin”, the dog walked in a 360-degree circle. Nominal task difficulty was varied while aiming to maintain relatively equal levels of functional task difficulty. Functional task difficulty was maintained by ensuring all dog participants had no prior experience with the tricks trained and conditions were maintained by training all dogs in the same environment.

Due to the unique nature of this thesis, we sought to provide enhanced measures of learning beyond task performance in the retention phase. In studies examining graduation rates of working dogs from training programs, researchers demonstrated distractibility to be a key factor affecting success (Bray et al., 2021). Along similar lines, Maejima et al. (2007) found that desire for work is a significant factor in the graduation of working dogs. Desire for work is defined as attending to the handler or the task at hand, while distractibility is characterized by the opposite (Jamieson et al., 2018; Maejima et al., 2007). These results align with two out of the three theories behind the CI effect, in which active reconstruction or elaboration of the tasks being learned is necessary for improved performance (Lee & Magill, 1985; Rendell et al., 2010; Shea & Morgan, 1979). The active nature of these theories suggests that attention to the tasks being trained is key during the learning process and distraction would have a potentially negative impact on performance. This is corroborated by Rochais et al. (2017) who demonstrated that horses with longer durations of distraction (in response to external stimuli) show less improvement on a visual discrimination task. Similarly in humans, Sanders and Baron (1975) reported that distraction significantly impairs human performance in complex learning tasks but not in simple ones. Finally, Inzlicht et al. (2018) affirmed both human and non-human animals will avoid tasks that require increased cognitive effort. Task avoidance in Inzlicht et al. (2018) is similar to disengagement from the handler, and opposite to desire for work. When examined in conjunction, the presented evidence suggests there could be a link between distraction and increased cognitive effort. It is unclear whether distraction is the cause of poor performance on complex tasks or a byproduct of increased cognitive load during complex tasks. However, research suggests that as task complexity increases distraction should become increasingly evident. The correlation between distraction and cognitive load led to our decision to measure disengagement from the handler as an indication of CI level in this learning paradigm.

In relation to the evidence presented above, we hypothesized the following:

1. During training, dogs in the random acquisition group would perform worse than dogs in the blocked acquisition group.
2. During a retention test, dogs who trained the three tasks in random order during acquisition would perform better than dogs who trained in blocked order during acquisition.
    a. Dogs who learned the three tasks in random order during acquisition and were tested in blocked order during the retention test would perform better than those tested in random order.
    b. Dogs who learned the three tasks in blocked order during acquisition and were tested in random order during the retention test would perform worse than those tested in blocked order.
3. During acquisition, dogs who learned the three tasks in random order would display higher levels of disengagement from the handler than the dogs who learned the tasks in blocked order.
    a. Dogs who trained the three tasks in blocked order during acquisition and were tested in random order during retention would show the highest levels of disengagement overall in the retention test.

b. Dogs who trained the three tasks in random order during acquisition and were tested in random order during retention would display moderate levels of disengagement in the retention test.

c. Dogs who trained the three tasks in blocked order during acquisition that were tested in blocked order during retention, as well as dogs who trained the three tasks in random order during acquisition that were tested in blocked order during retention, would show minimal levels of distraction as compared to the other groups in the retention test.

Methods

Participants and Facilities

Researchers enlisted 23 guardians and their pet dogs (*Canis lupus familiaris*) from the Thinking Dog Center database. The Thinking Dog Center (hereafter known as TDC) is located at 450 West 41st street in New York City. Interested guardians could sign their dog up to participate in a variety of in-person and online research studies via the link "tinyurl.com/ThinkingDogCenter." During the sign-up process, guardians answered questions related to their dog's health and listed any behavioral issues their dog may have. The TDC director, Dr. Sarah- Elizabeth Byosiere, deemed dogs fit for in-person or online studies based on the information provided. For this study dogs needed to be over the age of 4 months old, friendly, and food motivated. There were no breed restrictions for participating dogs. Guardians also had to commit to bringing their dog to the center for up to 1 hour a day on two consecutive weekdays. The dogs could have prior experience with basic manners training but to be considered for the study they were required not to have experience with three tricks, "chin rest," "platform," or "spin." Of the 23 dogs that were initially enrolled, only 17 dogs completed the study and were included in the final analyses. Six dogs were excluded due to behavioral issues (n = 4), failure to return on the second day of testing (n = 1) or becoming physically ill (because of the car ride beforehand) during testing (n = 1).

Guardian-dog teams selected for the study were offered their choice of 1-hour timeslots between the hours of 4-8 PM on consecutive Tuesdays and Wednesdays. Teams were required to return to the TDC at the same time of day on both days. Upon entrance to the testing facility, guardians and dogs were allowed a short acclimatization period in which dogs were given freedom to explore the space and guardians completed a consent form before testing proceeded. During this time, dogs could engage in play with toys set out, drink water freely available, eat treats given out by the handler, or explore the testing rooms as they wished. Guardians remained with their pet dogs throughout the entire study. If at any point, they felt their dog needed a break or wished to remove the dog from the environment they could. Of the 17 dogs included in the study, no guardians requested a break or the removal of their dog from the study. The dogs were also monitored by the handler for any signs of duress including but not limited to, panting, shaking, or inability to eat offered treats. The 17 dogs included in the statistical analyses did not display such signs of stress.

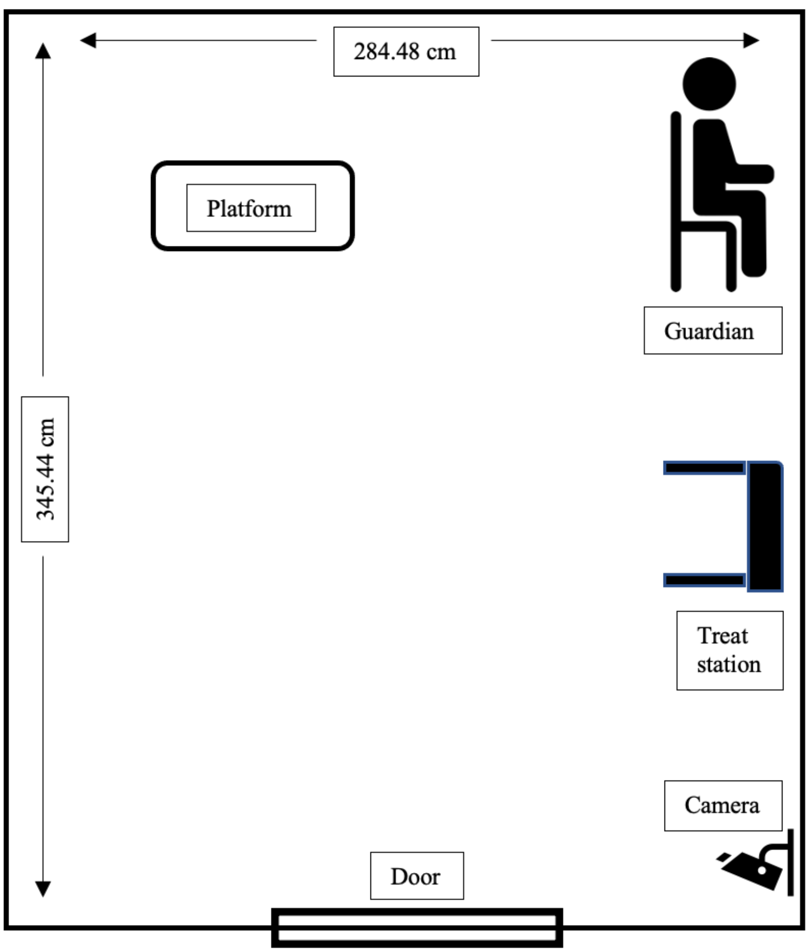
The dogs in this study were randomly assigned to four groups: blocked acquisition-blocked retention (n = 5); blocked acquisition - random retention (n = 4); random acquisition - random retention (n = 3); and random acquisition - blocked retention (n = 5). Due to the nature of the study and the possibility for a dog to be excluded during testing, there was an unequal number of dogs sorted into each group.

Procedure

The procedure for this experiment (SEB-Dog Tricks 5/22) was approved by Hunter CUNY's IACUC on June 12th, 2019. Acquisition and retention phases were conducted in a room measuring 345.44 cm x 284.48 cm. Guardians were instructed to sit on a chair in the far-right corner of the room opposite the entrance facing a wall (Figure 1). Throughout the visits, dogs were not movement restricted and could approach their guardians or move about the room as they pleased. Given this, guardians were asked not to speak to, engage with physically, or make eye contact with their dogs while testing was underway.

Figure 1

Depiction of Training and Testing Room with Measurements



Acquisition Phase

During the acquisition phase, the dogs were taught three tricks to be later tested on in the retention phase. The acquisition phase was broken down into three sessions with an approximate 1–5-minute break between sessions, depending on each dogs' individual needs for water and rest. Each session included 24 trials of tricks. A trial consisted of one presentation of the individual trick to be trained (e.g., asking the dog to perform one platform) (Table 1). Within each trial, the dogs were given two tries to complete a trick. A session for dogs in the blocked group consisted of training one trick repeatedly for 24 trials before moving on to the next trick in the following session (e.g., 24 trials of platform followed by 24 trials of spin followed by 24 trials of chin rest). The order in which tricks were presented over the three sessions in the blocked group was counterbalanced to account for order effects (Sheehe & Bross, 1961). A session for dogs in the random group included eight trials of each of the three tricks, however, the order the investigator presented the tricks was randomized (e.g., a trial of chin rest, followed by a trial of spin, followed by a trial of platform, and so on until each behavior had been presented an equal number of times, totaling 24 trials). This was repeated over the three sessions.

Retention Test

Dogs who successfully completed the acquisition phase, returned the next day to complete the retention test. The purpose of this test was to measure the amount of learning that took place during the acquisition phase. The retention test consisted of one session of 18 trials. Based on their initial group, participants were once again sorted into blocked and random retention sessions. The blocked group received 6 consecutive trials of each trick and the random group received 6 trials of each trick in a randomized order. The dogs had two tries to complete each trick within one trial.

Tricks

As described above, the first day of testing was called the acquisition phase, the second day was considered the retention phase. On each day the dogs were taught and tested on the tricks “chin rest,” “platform,” and “spin” by one handler. The author of this thesis trained all dogs who participated. The author is CPDT-KA certified and a professional dog trainer so was deemed an appropriate fit as the handler. Handler consistency in training all dogs also ensured consistency of training methods. The three tricks were taught using operant conditioning via the positive reinforcement-based training techniques, luring and shaping (Chiandetti et al., 2016). When training using luring, a handler holds a hidden piece of treat in between their fingers and uses it to move a dog around in space as they follow the hand (Chiandetti et al., 2016). With shaping, a handler works towards an end behavior by breaking it down into approximate steps and rewarding the dog as they obtain each consecutive step towards the end goal (Chiandetti et al., 2016). There were several levels a dog could reach within each trick (Table 1). A verbal secondary reinforcer “yes” was used before the delivery of the primary reinforcer (miscellaneous treats) throughout training as dogs accomplished different levels within the tricks (Chiandetti et al., 2016; Feng et al., 2016).

Chin Rest. For the chin rest trick, the handler started by placing one palm out at about chin height to the dog in front of her. The handler then used her other hand to lure the dog’s nose forward so their chin rested on the handler’s palm. At this point, the trick would be considered complete. The handler marked the position by saying “yes” and delivered a treat to the dog (Table 1).

Platform. The platform trick utilized a grey, plastic platform measuring 55.88cm L x 30.48cm W x 10.16cm H. The handler placed their lure in front of the dog’s nose and attempted to move the dog toward and onto the platform. The trick ended when the dog had both of their front paws on the platform at which point the handler marked “yes” and delivered a treat to the dog either in their mouth or on the floor nearby (Table 1).

Spin. For the spin trick, the handler brought their hand to the dog’s nose and lured their head around, so their body followed their head and the dog completed a 360 degree turn. The handler marked “yes” when the dog’s head had returned to starting position and delivered a treat to the dog (Table 1).

The chin rest trick was chosen to be the easy trick, the platform as the medium level difficulty trick, and the spin trick was chosen as the most difficult trick. Difficulty levels were determined by the proposed time to completion of the task and the number of body movements required to complete the goal action (Gorniak, 2019; Guadagnoli & Lee, 2004). Dogs were given two tries to reach the goal behavior for each trick. If the dog failed to follow the handler’s lure on the first try, the handler would pause and start the trick over. If the dog failed to follow the handler’s lure after the second try, the handler would reset the dog’s position by saying “find it” and tossing the treat in the lure hand on the ground for the dog to find.

Table 1

Ethogram for Coding Dog Behaviors During Acquisition and Retention Sessions

Behavior	Definition	Behavior Level	Level Definitions
Try 1	Used to define duration of first attempt at behavior. Beginning of first try marked by first movement of handler’s hands towards dog or at the first moment hand(s) are visible moving toward the dog. End of first try marked by removal of hands from in front of dog, if failed first try, or the sound of the word "yes" if successful.	NA	NA

Try 2	Used to define duration of second attempt of behavior. Beginning of second try marked by first movement of handler's hands towards dog or at the first moment hand(s) are visible moving toward the dog. End of second try signified by sound of the word "yes" or "find it."	NA	NA
Chin Rest	The easy trick. The goal of the trick is for the handler to lure dog's muzzle forward so that it is resting on the outstretched palm of handler's non-lure hand.	0 1 2	Missed first try Dog does not approach handler's palm Dog places muzzle in palm
Platform	The medium level trick. The goal of the platform trick was to lure the dog forward onto the platform, so the dog steps their two front paws on the platform.	0 1 2 3	Missed first try Dog does not follow handler's lure hand forward Dog places one paw on platform Dog places both eyes on platform
Spin	The hardest trick. The goal of the spin trick is to lure the dog 270 degrees or more around in a circle so that their head returns to starting position. Starting position is considered the point where the lure hand initially meets the dog's nose.	0 1 2 3 4	Missed first try Dog does not follow handler's lure with head turn Dog turns head 90 degrees Dog turns head 180 degrees Dog turns head 270 degrees or more
Disengage	Anytime the dog orients away from the handler (or her body parts) for 2 seconds or more unless leading into an OOV. Not including investigation of treats (on the ground or at the coding station).	NA	NA
Out of View (OOV)	Whenever the dogs' head and shoulders are out of view, so that one cannot see where the dog is orienting. Exceptions include when the dog is in handler's general vicinity (eg. under or behind her body, unless dog is clearly facing the opposite direction.)	NA	NA

Coding and Data Analyses

Acquisition and retention sessions were recorded on a Canon VIXIA HF R80 Camcorder located at the right-hand side of the doorway to the testing room (Figure 1). The camera view remained untouched throughout testing. Three coders – the researcher, and two naive Hunter Animal Behavior and Conservation Master's students – were responsible for coding all acquisition and retention videos. Approximately the same number of dogs' videos were randomly assigned to each coder. Coders used an ethogram to determine what level within each trick was reached and whether the dog took one or two tries to complete a behavior (Table 1). The latency to complete each trick, the duration of time dogs spent out of the camera view (OOV), and the duration of time dogs spent disengaged from the handler were also measured (Table 1). All coding was executed via the Behavioral Observation Research Interactive Software (BORIS), version 7.9.19 (Friard & Gamba, 2016).

Reliability

Inter-rater reliability (IRR) was determined based on four dogs' acquisition and retention videos (23.5% of the dogs). Gwet's AC1 was performed using AgreeStat 360 a free internet platform (AgreeStat 360, n.d). AC1 was used with ordinal behavior and lure scores as it is less sensitive to trait prevalence and trait bias than Fleiss's Kappa (Gwet, 2008). When there is high agreement between raters but low variance, especially for a small sample size, Fleiss's Kappa tends to be paradoxically low (Hoek & Scholman, 2017; Xie, 2013). Gwet's AC1 is less affected by skewed distributions and therefore better reflected the level of agreement between raters in this study (Dettori & Norvell, 2020).

Gwet's AC1 for behavior score consistency across all dogs and between the three coders was 0.96 (SE 0.01), which is considered "very good" in the benchmark range (Dettori & Norvell, 2020). For agreement on disengagement (D) and out of view (OOV) instances, Gwet's AC1 was 0.52 (SE 0.06). The AC1 for D/OOV instances is in the moderate range (Dettori & Norvell, 2020).

To assess agreement between coders on duration of D and OOV instances, the intraclass correlation coefficient (ICC) was used. The ICC was the suggested test for continuous data, coded by three or more judges (Koo & Li, 2016; Perinetti, 2018). A two-way, mixed-effects model was used based on single ratings to assess absolute agreement on continuous measures of duration. For duration of behavior scores the ICC was 0.79 (95% CI, 0.73 - 0.84) and for duration of D/OOV scores the ICC was 0.60 (95% CI, 0.49 - 0.70). The ICC for duration of behavior scores (0.79) was considered "good" and for duration of D/OOV scores (0.60) the ICC fell in the "fair" range (Koo & Li, 2016).

Statistical Analyses

Composite trick scores were created by summing the trick levels coded for each individual dog and averaging the sum over the number of tries completed. The latency to complete each try of the tricks was also averaged over the number of tries per trick. The durations of OOV instances were summed to find the total time a dog spent OOV per acquisition and retention phase. Percentage of time spent disengaged was calculated as such:

$$\%\,Time\ Spent\ Disengaged = \left(\frac{Total\ Disengagement\ Time}{(Total\ Video\ Time - Total\ Time\ Dog\ Spent\ OOV}\right) \times\ 100$$

The number of trials before reaching maintenance of the goal behavior was deduced for each trick. Maintenance of a trick was reached on the third performance of the highest behavior level within chin rest, platform, or spin (Table 1). For each trick, the lowest possible trick maintenance score was therefore three. The first occurrence of trick maintenance for chin rest, platform, or spin was recorded for each acquisition and retention phase. The total number of second tries for each trick was recorded per session as well.

All statistical analyses were conducted using IBM SPSS Statistics 27. Variables from the acquisition and retention data were assessed for normality using skewness and kurtosis values. Values between ± 2 were considered to be normally distributed (George & Mallery, 2009). Approximately half of the variables had skewness and kurtosis values that fell outside of the ± 2 range, as a result, non-parametric tests were explored as a suitable option for analyzing the dataset (Table 2). The use of non-parametric tests, in this case, was supported by evidence suggesting that when assumptions of parametric tests are not met and the sample size falls between 8 and 24 participants non-parametric tests are suitable (Dwivedi et al., 2017). Mann-Whitney U tests were run to assess if there were any differences between the blocked and random groups in acquisition and retention. Kruskal-Wallis tests were run to assess differences between the retention groups with respect to the acquisition groups.

Table 2

Skewness and Kurtosis Values for Acquisition and Retention Variables

Acquisition Variables	Variable	Skewness SD ± 0.64	Kurtosis SD ± 1.23
Chin	Second Tries	1.15	0.55
	Duration	0.56	-0.55
	Score	-0.68	-0.74
	Trials to Maintenance	2.16	3.44
Platform	Second Tries	2.61	6.87
	Duration	2.15	4.83
	Score	-2.01	3.09
	Trials to Maintenance	1.31	3.02
Spin	Second Tries	-0.68	-0.89
	Duration	0.15	-0.76
	Score	-0.96	0.44
	Trials to Maintenance	1.05	1.33

Disengagement	Percent Disengaged	1.32	0.75
Retention Variables	Variable	Skewness SD ± 0.62	Kurtosis SD ± 1.19
Chin	Second Tries	2.05	3.71
	Duration	0.86	-0.02
	Score	-1.86	2.52
	Trials to Maintenance	3.28	11.03
Platform	Second Tries	3.61	13.00
	Duration	1.03	0.63
	Score	-3.61	13.00
	Trials to Maintenance	3.61	13.00
Spin	Second Tries	0.93	-0.98
	Duration	1.45	3.59
	Score	-0.83	-1.24
	Trials to Maintenance	2.68	6.97
Disengagement	Percent Disengaged	2.81	8.49

Results

Hypothesis 1

Mann-Whitney U tests were run to determine if there were differences between dogs in the blocked and random acquisition groups. Distributions for all variables were visually inspected and assessed to be similar. There were no significant differences in second tries, duration, score, or trials to maintenance variables across the chin rest, platform, and spin tricks for blocked and random acquisition groups, $p > .05$ (results presented in Table 3 due to the large number of variables measured).

Table 3

Summary of Mann-Whitney U Test Results for Variables Measuring Learning and Disengagement Across Chin Rest, Platform, and Spin in the Acquisition Session.

Variable	Median	U	z	Sig.[a]
Chin				
Second Tries	Blocked (N=9) = 2.00 Random (N=8) = 1.50	31.00	-0.49	.67

Duration	Blocked (N=9) = 2.01 Random (N=8) = 2.69	54.00	1.73	.09
Score	Blocked (N=9) = 1.85 Random (N=8) = 1.86	40.00	0.39	.74
Trials to Maintenance	Blocked (N=9) = 3.00 Random (N=8) = 3.00	37.00	0.17	1.00
Platform				
Second Tries	Blocked (N=9) = 1.00 Random (N=8) = 0.50	36.00	0.00	1.00
Duration	Blocked (N=9) = 2.64 Random (N=8) = 3.51	44.00	0.77	.48
Score	Blocked (N=9) = 2.84 Random (N=8) = 2.84	31.50	-0.44	.67
Trials to Maintenance	Blocked (N=8) = 4.50 Random (N=6) = 4.50	24.00	0.00	1.00
Spin				
Second Tries	Blocked (N=9) = 4.00 Random (N=8) = 2.00	23.50	-1.20	.24
Duration	Blocked (N=9) = 3.36 Random (N=8) = 3.48	41.00	0.48	.67
Score	Blocked (N=9) = 3.26 Random (N=8) = 3.48	47.00	1.06	.32
Trials to Maintenance	Blocked (N=6) = 8.00 Random (N=8) = 8.00	25.50	0.20	.85
Disengagement				
Percent Disengaged	Blocked (N=9) = 1.43 Random (N=8) = 3.08	52.00	1.54	.14

a. Exact significance is displayed for this test.

Hypothesis 2

To determine if there were any significant differences between blocked and random retention groups, Mann-Whitney U tests were performed for learning variables across chin rest, platform, and spin. The distributions for all variables were visually assessed and found to be similar. No significant differences between blocked and random retention groups were found for second tries, no lure, duration, score, or trials to maintenance variables, $p > .05$ (results for each test displayed in Table 4 due to the large number of variables measured).

Table 4

Summary of Mann-Whitney U Test Results for Variables Measuring Learning and Disengagement Across Chin Rest, Platform, and Spin in the Retention Session.

Variable	Median	U	z	Sig.[a]
Chin				
Second Tries	Blocked (N=10) = 0.00 Random (N=7) = 0.00	37.50	0.30	.81
Duration	Blocked (N=10) = 1.84 Random (N=7) = 2.17	55.00	1.95	.06
Score	Blocked (N=10) = 2.00 Random (N=7) = 2.00	32.50	-0.30	.81
Trials to Maintenance	Blocked (N=10) = 3.00 Random (N=6) = 3.00	32.50	0.47	.79
Platform				
Second Tries	Blocked (N=10) = 0.00 Random (N=7) = 4.00	24.50	-1.55	.32
Duration	Blocked (N=10) = 1.91 Random (N=7) = 3.89	50.00	1.46	.16
Score	Blocked (N=10) = 3.00 Random (N=7) = 3.00	34.50	-0.06	.96
Trials to Maintenance	Blocked (N=9) = 3.00 Random (N=5) = 3.00	17.50	-1.10	.52
Spin				
Second Tries	Blocked (N=10) = 0.50 Random (N=7) = 0.00	28.00	-0.77	.54
Duration	Blocked (N=10) = 2.80 Random (N=7) = 3.23	42.00	0.68	.54
Score	Blocked (N=10) = 3.71 Random (N=7) = 4.00	40.00	0.53	.67
Trials to Maintenance	Blocked (N=9) = 3.00 Random (N=7) = 3.00	33.50	0.28	.84
Disengagement				

Percent Disengaged	Blocked (N=10) = 1.68 Random (N=7) = 0.25	24.50	-1.05	.32

a. Exact significance is displayed for this test.

Hypothesis 2a. and Hypothesis 2b

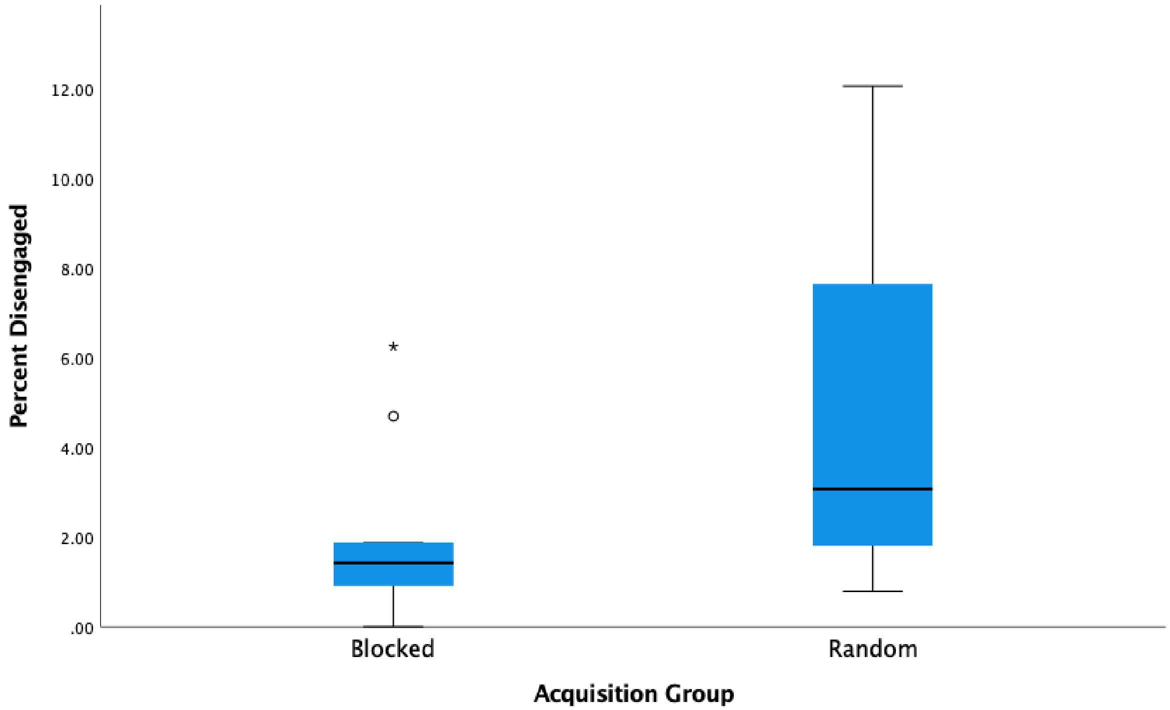
Kruskal-Wallis H tests were run to determine if there were any differences between the retention groups with respect to acquisition group for learning variables across chin rest, platform, and spin. The boxplot distributions for all variables were visually assessed and found to be similar. No significant differences between blocked-blocked, random-blocked, blocked-random, and random-random retention groups were found for second tries, no lure, duration, score, or trials to maintenance variables, $p > .05$ (results for each test displayed in Table 5 due to the large number of variables measured).

Hypothesis 3

Mann-Whitney U tests were run to determine if there was a difference between blocked and random groups on percentage of time spent disengaged from the handler in both the acquisition and retention sessions. Distributions were visually assessed to be similar prior to completing the tests. As hypothesized, in acquisition median percent disengaged for the random group (Mdn = 3.08) was higher than median percent disengaged for the blocked group (Mdn = 1.43) however this was not statistically significant, $U = 52$, $z = 1.54$, $p = .14$ (Figure 2; Table 3). There was no significant difference between the blocked (Mdn = 1.68) and random (Mdn = 0.25) retention groups for percent of time spent disengaged, $U = 24.50$, $z = -1.05$, $p = .32$ (Table 4).

Figure 2

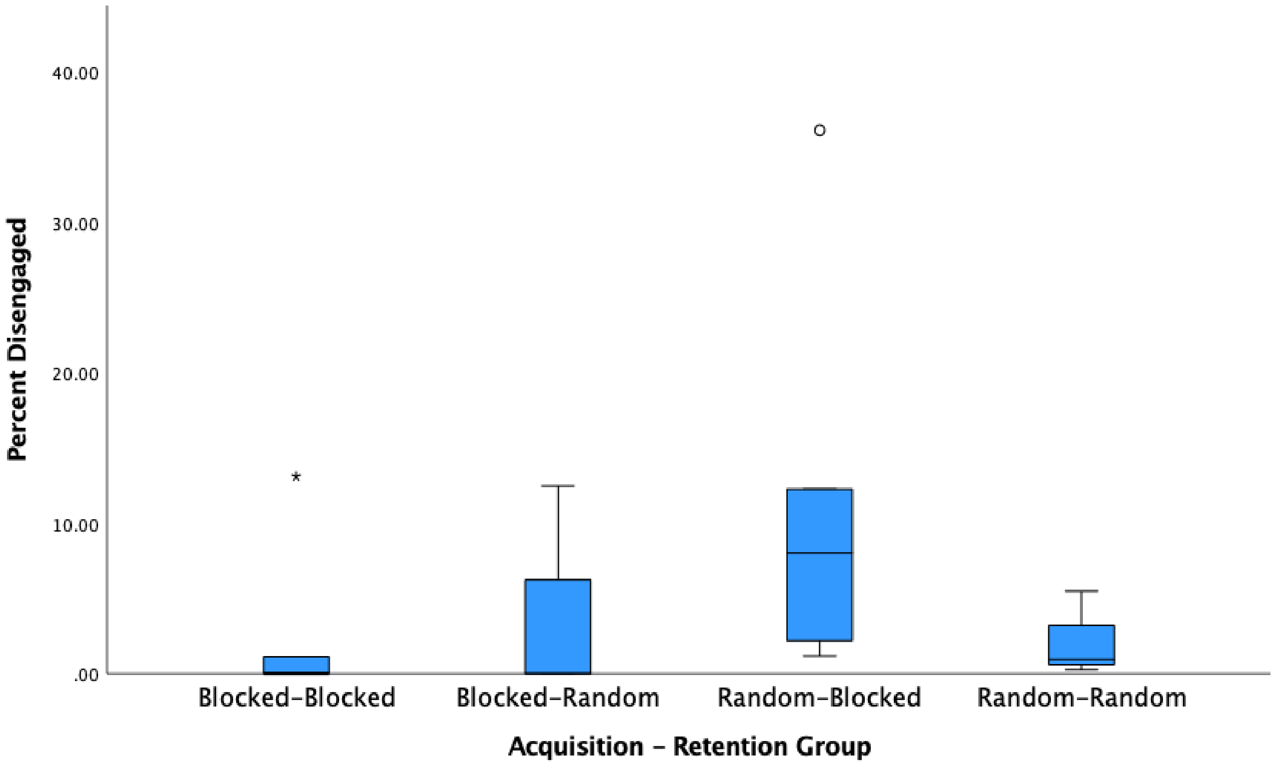
Percent Disengaged over Acquisition



Note. The graph displays percent disengaged for blocked and random acquisition groups. Circles on the graph represent outliers and stars represent extreme outliers.

Figure 3

Percent Disengaged over the Retention Sessions



Note. The graph displays the percent disengaged in retention groups with respect to acquisition group. Circles on the graph represent outliers and stars represent extreme outliers.

Hypothesis 3a., Hypothesis 3b., and Hypothesis 3c

To determine if there were any differences in percent disengaged during retention between blocked-blocked, random-blocked, blocked-random, and random-random retention groups Kruskal-Wallis H tests were run. Boxplot distributions for all variables were determined to be similar by visual assessment. Contrary to hypotheses 3a-3c, the blocked-random group did not display the highest median percent disengaged in retention (Mdn = 0.00) (Figure 3). The blocked-random group displayed similarly low levels of disengagement during retention as the blocked-blocked (Mdn = 0.00) and random-random (Mdn = 0.93) groups. The random-blocked group spent the most time disengaged during retention (Mdn = 8.02). The differences for percent disengaged across blocked-blocked, random-blocked, blocked-random, and random-random retention groups were not statistically significant, $p > .05$ (results for each test displayed in Table 5 due to the large number of variables measured).

Table 5

Summary of Kruskal-Wallis Test Results, Grouped by Acquisition and Retention Condition, for Variables Measuring Learning and Disengagement Across Chin Rest, Platform, and Spin in the Retention Session.

Variables	Median	$\chi^2(3)$	Sig.[a]
Chin			
Second Tries	Blocked-Blocked (N=5) = 1.00 Blocked-Random (N=4) = 0.00 Random-Blocked (N=5) = 0.00 Random-Random (N=3) = 0.00	3.30	.35
Duration	Blocked-Blocked (N=5) = 1.97 Blocked-Random (N=4) = 2.04 Random-Blocked (N=5) = 1.75 Random-Random (N=3) = 3.01	5.21	.16
Score	Blocked-Blocked (N=5) = 1.71 Blocked-Random (N=4) = 2.00 Random-Blocked (N=5) = 2.00 Random-Random (N=3) = 2.00	3.30	.35
Trials to Maintenance	Blocked-Blocked (N=5) = 3.00 Blocked-Random (N=4) = 3.00 Random-Blocked (N=5) = 3.00 Random-Random (N=2) = 3.00	1.78	.62
Platform			
Second Tries	Blocked-Blocked (N=5) = 0.00 Blocked-Random (N=4) = 3.50 Random-Blocked (N=5) = 0.00 Random-Random (N=3) = 4.00	3.05	.38
Duration	Blocked-Blocked (N=5) = 2.05 Blocked-Random (N=4) = 2.56 Random-Blocked (N=5) = 1.51 Random-Random (N=3) = 4.00	5.40	.15

Score	Blocked-Blocked (N=5) = 3.00 Blocked-Random (N=4) = 3.00 Random-Blocked (N=5) = 3.00 Random-Random (N=3) = 3.00	0.86	.84
Trials to Maintenance	Blocked-Blocked (N=5) = 3.00 Blocked-Random (N=3) = 3.00 Random-Blocked (N=4) = 3.00 Random-Random (N=2) = 3.00	1.25	.74
Spin			
Second Tries	Blocked-Blocked (N=5) = 1.00 Blocked-Random (N=4) = 0.00 Random-Blocked (N=5) = 0.00 Random-Random (N=3) = 0.00	1.85	.61
Duration	Blocked-Blocked (N=5) = 2.42 Blocked-Random (N=4) = 3.00 Random-Blocked (N=5) = 2.90 Random-Random (N=3) = 3.92	2.76	.43
Score	Blocked-Blocked (N=5) = 3.14 Blocked-Random (N=4) = 4.00 Random-Blocked (N=5) = 4.00 Random-Random (N=3) = 3.67	2.65	.45
Trials to Maintenance	Blocked-Blocked (N=4) = 3.50 Blocked-Random (N=4) = 3.00 Random-Blocked (N=5) = 3.00 Random-Random (N=3) = 3.00	2.80	.42
Disengagement			
Percent Disengaged	Blocked-Blocked (N=5) = 0.00 Blocked-Random (N=4) = 0.00 Random-Blocked (N=5) = 8.02 Random-Random (N=3) = 0.93	5.15	.16

a. Asymptotic significance is displayed.

Discussion

The present study aims to apply the highly explored contextual interference theory to dog training practices. The results are not indicative of the CI effect in dog learning. In the acquisition phase, if the CI effect were present, we would expect to see decreased learning scores and increased disengagement scores for the random group, as compared to the blocked group.

This was not evident in the learning variables. While percent disengaged was larger for the random group than the blocked group, this result was not statistically significant. In the retention phase, contrary to our expectations, the random-blocked and random-random (acquisition-retention) groups did not score higher on the learning variables than the blocked-blocked or the blocked-random groups. The expectation that in retention, random acquisition groups would show less disengagement from the handler than blocked acquisition groups was also not met.

While our findings are not statistically significant, the lack of strong evidence for the CI effect on learning in dogs aligns with several studies that failed to show the CI effect in human learning paradigms (Barreiros et al., 2007; Farrow & Buszard, 2017; Wulf & Shea, 2002). As previously mentioned, there are a few arguments crediting specific study design aspects for this failure in human CI research. These are the use of laboratory vs. applied design, the regulation of task complexity, and inclusion of propulsive vs. non propulsive actions which are discussed in detail below.

Suggestions for Study Design Improvements

First, this research was conducted in a “laboratory” setting. The TDC is set up to be as comfortable for dogs as possible and each dog participant was given time to adjust to the new setting. However, this does not change the fact that many companion dogs are probably trained on novel skills first in their own homes rather than in a strange environment. Ultimately, they are expected to perform these skills in a public setting. This is not unlike many athletes who learn skills in a practice setting and then are required to perform said skills in a competitive setting. In a review of CI in applied settings, Farrow and Buszard (2017) recommend using transfer tests over retention tests when examining CI in humans. In particular, the authors suggest the use of competitive settings for transfer tests as this most closely mimics the learning requirements placed on professional athletes. Two possible applications of this recommendation to future CI studies in dogs would be 1) conducting acquisition training in companion dogs' homes and transfer testing in a novel environment like the TDC or 2) conducting acquisition training in an environment such as the TDC, after a longer habituation session, and conducting transfer tests in real-life settings such as on the streets/parks of New York City. The first option seems the most valid as conditions in city streets/parks can be highly variable.

The second applicable human CI study design topic is the regulation of skill complexity. Functional task difficulty was regulated by ensuring that all training and testing took place in the same environment and that no dog participants had prior experience with the three tricks, chin rest, platform, and spin. However, the degree of experience each dog previously had with marker training was not controlled for. This may have implications for the efficacy of marker usage and therefore the learning taking place within the CI paradigm (Chiandetti et al., 2016; Feng et al., 2016). Individual dog temperaments may also play a role in the difficulty of learning new skills in an environment such as the TDC, especially if certain dogs are prone to fear or stress in new environments. In fact, increased fear has been negatively correlated with trainability in dogs (Hare et al., 2018). In the Demant et al. (2011) study that was discussed in the introduction, the beagle participants were sorted into groups based on their personalities. Future applications of the CI effect in dogs should control for level of marker training experience and temperament differences which may affect learning ability.

The last aspect of human CI designs that may have played a role in this study is the investigation of propulsive actions. Propulsive actions are defined by Barreiros et al. (2007) as actions related to throwing and kicking, in other words actions that require the application of force on another object. The chin rest and spin tricks did not require propulsive actions as far as this author can tell. The platform trick included the element of stepping onto the platform. This could be seen as a propulsive action as the dog needed to push off the platform with one paw to place the second paw on the platform surface. The degree to which propulsive actions affect CI findings is still under debate. As such, the potential propulsive aspect of the platform trick may be an impediment to finding the CI effect in dog training.

Beyond the author's control, the experimental phase of this study took place in late 2019 and early 2020. Therefore, the COVID-19 pandemic did affect the ability to train and test dog participants. Originally, the study was set-up to assess a minimum of 24 total participants. In the end, only 17 dogs were included in the study. Working with a larger sample size could lessen the effects of the aforementioned issues applicable to contextual interference research, by providing greater variability of dog temperament and training experience within each group. Similarly, a larger sample size could increase the statistical power of non-parametric and parametric analyses when examining differences between the CI groups (Dwivedi et al., 2017; Janušonis, 2009).

The CI Paradigm as a Framework for Future Research

The importance of research directing training practices in both the companion and working dog industries cannot be understated. In a recent review of working dog training practices, Hall et al. (2021) highlight that despite the widespread use of classical and operant conditioning techniques in working dog task training, there is little research investigating the efficiency of these training techniques and their effect on dog performance. Much of the recent literature examining dog training efficacy centers around the use of operant conditioning to train one skill at a time (Chiandetti et al., 2016; Demant et al., 2011; Feng et al., 2016; Meyer & Ladewig, 2008). These four studies highlight the disconnect between research in companion and working dog training where dogs are often trained on several skills at a time within a training session. The current thesis offers a potential solution to this dilemma by providing a training paradigm for assessing the efficacy of focusing on one skill compared to several skills per training session. Though our results are not definitive as to which practice format (blocked or random) is more beneficial for the performance of these skills, we provide an example paradigm which may be built upon further to support one practice format over the other.

Within these four studies, there are also variations on length of training session. In Meyer and Ladewig (2008) sessions consisted of 15 trials, in Demant et al. (2011) sessions were six trials presented once/three times in a row, in Feng et al. (2018) sessions were 15-20 mins, and in Chiandetti et al. (2016) training sessions were conducted three consecutive times for an unknown length in trials or minutes. In our CI paradigm, dogs completed three consecutive sessions of 24 trials. Given that, Demant et al. (2011) demonstrate support for shorter training periods (one over three sessions in a row) it’s possible that all the paradigms mentioned, including this thesis work, used session lengths that are too long and may have impacted dogs’ performance negatively. However, support for shorter training periods was shown when dogs were trained on only one trick at a time. The current thesis paradigm provides a foundational model for this approach with several skills at a time. In one session of the random acquisition phase, dogs performed eight trials of each of the three tricks, resulting in 24 trials total. Rather than asking dogs to participate in three consecutive sessions, these three sessions could be spread out over the course of three weeks. It’s possible that by doing so, as did Demant et al (2011), performance may increase. Future research may continue to build off our CI paradigm in this way to understand the effect of session length on both single and multiple trick training.

This thesis and its implications for future research, informing best practices for training several skills and training session length, can significantly impact the companion and working dog training industries. One of the most popular questions many companion dog guardians ask is: What should training my dog look like (Coren, 2013; Todd, 2012)? They are most often referring to how frequently they should be training their dogs on a daily or weekly basis and for how long. This question also encompasses whether they should focus on one skill at a time or several skills within one training session. Up until now, there was only research to support short training sessions given once weekly. While our CI paradigm did not provide significant results suggesting that either blocked or random format for training multiple skills is better than the other, this thesis provides a framework for answering that question.

One area of the study that we hope to see more of in future animal learning research is a secondary measure such as disengagement. Though the difference was not significant in acquisition, the random group seemed to be trending towards showing more disengagement from the handler than the blocked group. Anecdotally, this could be support for increased cognitive load resulting from learning skills in a random format. This is complemented by the three theories behind the CI effect in humans, suggesting that increased cognitive effort is necessary for the CI effect to take place (Farrow & Buszard, 2017). While random training did not seem to benefit dogs in retention over blocked training, our retention results may have been affected by the small dataset (mentioned previously). Still, these results are encouraging and could certainly be used to promote disengagement as a measure of cognitive load in the future. This would be beneficial because, as evidence in humans suggests, once someone has reached a high level of cognitive load they may no longer be experiencing the beneficial effect of CI and therefore learning performance could be impacted (Farrow & Buszard, 2017). In particular, this could have significant implications for working dogs who typically face physically and mentally grueling training and performance conditions (Hall et al., 2021) If handlers had prior knowledge of what degree of disengagement suggests no CI effect, some CI effect, or too much CI effect on learning, they could alter their training practices to ensure better performance in dogs. This might mean cutting certain training sessions short if they see too much disengagement, while increasing the difficulty of other training sessions if they are seeing little to no disengagement. In this way, measuring cognitive load could help inform best training practices.

In conclusion, as the dog training industry continues to grow, so will the importance of research instructing both the companion and working dog fields on best training practices. It should be stressed that the current thesis paves the way for future attempts at dissecting the efficacy of CI on learning in dogs. As demonstrated, this research could support both the companion dog and service dog industries in increasing their success rates when it comes to behavior change.

References

Arden, R., Bensky, M. K., & Adams, M. J. (2016). A review of cognitive abilities in dogs, 1911 through 2016: More individual differences, please! *Current Directions in Psychological Science*, *25*(5), 307–312. https://doi.org/10.1177/0963721416667718

Barreiros, J., Figueiredo, T., & Godinho, M. (2007). The contextual interference effect in applied settings. *European Physical Education Review*, *13*(2), 195–208. https://doi.org/10.1177/1356336X07076876

Battig, W. F. (1972). *Interference during learning as a sources of facilitation in subsequent retention and transfer*. https://eric.ed.gov/?id=ED062661

Bennett, P., Cooper, N., Rohlf, V., & Mornement, K. (2007). Factors influencing owner satisfaction with companion-dog-training facilities. *Journal of Applied Animal Welfare Science*, *10*(3), 217–241. https://doi.org/10.1080/10888700701353626

Bogaerts, E., Moons, C. P. H., Van Nieuwerburgh, F., Peelman, L., Saunders, J. H., & Broeckx, B. J. G. (2019). Rejections in an non-purpose bred assistance dog population: Reasons, consequences and methods for screening. *PLoS ONE*, *14*(6), e0218339. https://doi.org/10.1371/journal.pone.0218339

Brady, F. (1997). Contextual interference and teaching golf skills. *Perceptual and Motor Skills*, *84*(1), 347–350. https://doi.org/10.2466/pms.1997.84.1.347

Brady, K., Cracknell, N., Zulch, H., & Mills, D. S. (2018). A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Frontiers in Veterinary Science*, *5*(103), 1–10. https://doi.org/10.3389/fvets.2018.00103

Bray, E. E., Levy, K. M., Kennedy, B. S., Duffy, D. L., Serpell, J. A., & MacLean, E. L. (2019). Predictive models of assistance dog training outcomes using the canine behavioral assessment and research questionnaire and a standardized temperament evaluation. *Frontiers in Veterinary Science*, *6*(49), 1–11. https://doi.org/10.3389/fvets.2019.00049

Bray, E. E., Otto, C. M., Udell, M. A. R., Hall, N. J., Johnston, A. M., & MacLean, E. L. (2021). Enhancing the selection and performance of working dogs. *Frontiers in Veterinary Science*, *8*(644431), 1–21. https://doi.org/10.3389/fvets.2021.644431

Chiandetti, C., Avella, S., Fongaro, E., & Cerri, F. (2016). Can clicker training facilitate conditioning in dogs? *Applied Animal Behaviour Science*, *184*(2016), 109–116. https://doi.org/10.1016/j.applanim.2016.08.006

China, L., Mills, D. S., & Cooper, J. J. (2020). Efficacy of dog training with and without remote electronic collars vs. a focus on positive reinforcement. *Frontiers in Veterinary Science*, *7*(508), 1–11. https://doi.org/10.3389/fvets.2020.00508

Cobb, M., Branson, N., McGreevy, P., Lill, A., & Bennett, P. (2015). The advent of canine performance science: Offering a sustainable future for working dogs. *Behavioural Processes*, *110*, 96–104. https://doi.org/10.1016/j.beproc.2014.10.012

Coren, S. (2013, September 23). *What type of training schedule works best for dogs?* Psychology Today. https://www.psychologytoday.com/us/blog/canine-corner/201309/what-type-training-schedule-works-best-dogs

Deldalle, S., & Gaunet, F. (2014). Effects of 2 training methods on stress-related behaviors of the dog (Canis familiaris) and on the dog–owner relationship. *Journal of Veterinary Behavior*, *9*(2), 58–65. https://doi.org/10.1016/j.jveb.2013.11.004

Demant, H., Ladewig, J., Balsby, T. J. S., & Dabelsteen, T. (2011). The effect of frequency and duration of training sessions on acquisition and long-term memory in dogs. *Applied Animal Behaviour Science*, *133*(3), 228–234. https://doi.org/10.1016/j.applanim.2011.05.010

Dettori, J. R., & Norvell, D. C. (2020). Kappa and beyond: Is there agreement? *Global Spine Journal*, *10*(4), 499–501. https://doi.org/10.1177/2192568220911648

Dorey, N. R., & Cox, D. J. (2018). Function matters: A review of terminological differences in applied and basic clicker training research. *PeerJ*, *6*(5621), 1–17. http://dx.doi.org.proxy.wexler.hunter.cuny.edu/10.7717/peerj.5621

Dwivedi, A. K., Mallawaarachchi, I., & Alvarado, L. A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in Medicine*, *36*(14), 2187–2205. https://doi.org/10.1002/sim.7263

Farrow, D., & Buszard, T. (2017). Exploring the applicability of the contextual interference effect in sports practice. *Progress in Brain Research*, *234*(2017), 69–83. https://doi.org/10.1016/bs.pbr.2017.07.002

Feng, L. C., Hodgens, N. H., Woodhead, J. K., Howell, T. J., & Bennett, P. C. (2018). Is clicker training (Clicker + food) better than food-only training for novice companion dogs and their owners? *Applied Animal Behaviour Science*, *204*(2018), 81–93. https://doi.org/10.1016/j.applanim.2018.04.015

Feng, L. C., Howell, T. J., & Bennett, P. C. (2016). How clicker training works: Comparing Reinforcing, Marking, and Bridging hypotheses. *Applied Animal Behaviour Science*, *181*(2016), 34–40. https://doi.org/10.1016/j.applanim.2016.05.012

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Fugazza, C., & Miklósi, Á. (2014). Should old dog trainers learn new tricks? The efficiency of the Do as I do method and shaping/clicker training method to train dogs. *Applied Animal Behaviour Science*, *153*(2014), 53–61. https://doi.org/10.1016/j.applanim.2014.01.009

Fugazza, C., & Miklósi, Á. (2015). Social learning in dog training: The effectiveness of the Do as I do method compared to shaping/clicker training. *Applied Animal Behaviour Science*, *171*(2015), 146–151. https://doi.org/10.1016/j.applanim.2015.08.033

George, D., & Mallery, P. (2009). *SPSS for Windows step by step: A simple study guide and reference, 17.0 update* (10th ed.). Allyn & Bacon, Inc.

Gilchrist, R. J., Gunter, L. M., Anderson, S. F., & Wynne, C. D. L. (2021). The click is not the trick: The efficacy of clickers and other reinforcement methods in training naïve dogs to perform new tasks. *PeerJ*, *9*(10881), 1–24. https://doi.org/10.7717/peerj.10881

Goode, S., & Magill, R. A. (1986). Contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport*, *57*(4), 308–314. https://doi.org/10.1080/02701367.1986.10608091

Gorniak, S. L. (2019). The relationship between task difficulty and motor performance complexity. *Attention, Perception, & Psychophysics*, *81*(1), 12–19. https://doi.org/10.3758/s13414-018-1634-4

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, *36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Gwet, K. L. (n.d.). *AgreeStat 360: Cloud-based inter-rater reliability analysis, Cohen's kappa, Gwet's AC1/AC2, Krippendorff's alpha, Brennan-Prediger, Fleiss generalized kappa, intraclass correlation coefficients*. Retrieved June 22, 2021, from https://agreestat360.com/

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. https://doi.org/10.1348/000711006X126600

Hall, N. J., Johnston, A. M., Bray, E. E., Otto, C. M., MacLean, E. L., & Udell, M. A. R. (2021). Working dog training for the twenty-first century. *Frontiers in Veterinary Science*, *8*(646022), 1–18. https://doi.org/10.3389/fvets.2021.646022

Hare, E., Kelsey, K. M., Serpell, J. A., & Otto, C. M. (2018). Behavior differences between search-and-rescue and pet dogs. *Frontiers in Veterinary Science*, *5*(118), 1–5. https://doi.org/10.3389/fvets.2018.00118

Hoek, J., & Scholman, M. (2017). Evaluating discourse annotation: Some recent insights and new approaches. *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-13)*. https://aclanthology.org/W17-7401

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Jamieson, L. T. J., Baxter, G. S., & Murray, P. J. (2018). You are not my handler! Impact of changing handlers on dogs' behaviours and detection performance. *Animals*, *8*(176), 1–11. https://doi.org/10.3390/ani8100176

Janušonis, S. (2009). Comparing two small samples with an unstable, treatment-independent baseline. *Journal of Neuroscience Methods*, *179*(2), 173–178. https://doi.org/10.1016/j.jneumeth.2009.01.017

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kwan, J. Y., & Bain, M. J. (2013). Owner attachment and problem behaviors related to relinquishment and training techniques of dogs. *Journal of Applied Animal Welfare Science*, *16*(2), 168–183. https://doi.org/10.1080/10888705.2013.768923

Lea, S. E. G., & Osthaus, B. (2018). In what sense are dogs special? Canine cognition in comparative context. *Learning & Behavior*, *46*(4), 335–363. https://doi.org/10.3758/s13420-018-0349-7

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In *Advances in Psychology* (Vol. 27, pp. 3–22). North-Holland. https://doi.org/10.1016/S0166-4115(08)62528-5

Lin, C.-H. (Janice), Yang, H.-C., Knowlton, B. J., Wu, A. D., Iacoboni, M., Ye, Y.-L., Huang, S.-L., & Chiang, M.-C. (2018). Contextual interference enhances motor learning through increased resting brain connectivity during memory consolidation. *NeuroImage*, *181*(2018), 1–15. https://doi.org/10.1016/j.neuroimage.2018.06.081

MacLean, E. L., & Hare, B. (2018). Enhanced selection of assistance and explosive detection dogs using cognitive measures. *Frontiers in Veterinary Science*, *5*(236), 1–14. https://doi.org/10.3389/fvets.2018.00236

MacLean, E. L., Herrmann, E., Suchindran, S., & Hare, B. (2017). Individual differences in cooperative communicative skills are more similar between dogs and humans than chimpanzees. *Animal Behaviour*, *126*(2017), 41–51. https://doi.org/10.1016/j.anbehav.2017.01.005

Maejima, M., Inoue-Murayama, M., Tonosaki, K., Matsuura, N., Kato, S., Saito, Y., Weiss, A., Murayama, Y., & Ito, S. (2007). Traits and genotypes may predict the successful training of drug detection dogs. *Applied Animal Behaviour Science*, *107*(3), 287–298. https://doi.org/10.1016/j.applanim.2006.10.005

Meyer, I., & Ladewig, J. (2008). The relationship between number of training sessions per week and learning in dogs. *Applied Animal Behaviour Science*, *111*(3–4), 311–320. https://doi.org/10.1016/j.applanim.2007.06.016

Ollis, S., Button, C., & Fairweather, M. (2005). The influence of professional expertise and task complexity upon the potency of the contextual interference effect. *Acta Psychologica*, *118*(3), 229–244. https://doi.org/10.1016/j.actpsy.2004.08.003

Otto, C. M., Sell, T. K., Veenema, T., Hosangadi, D., Vahey, R. A., Connell, N. D., & Privor-Dumm, L. (2021). The promise of disease detection dogs in pandemic response: Lessons learned from COVID-19. *Disaster Medicine and Public Health Preparedness*, 1–19. https://doi.org/10.1017/dmp.2021.183

Pauwels, L., Swinnen, S., & Beets, I. (2014). Contextual interference in complex bimanual skill learning leads to better skill persistence. *PLoS One*, *9*(6), 1–10. https://doi.org/10.1371/journal.pone.0100906

Perinetti, G. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South European Journal of Orthodontics and Dentofacial Research*, *5*(1), 3–5. https://doi.org/10.5937/sejodr5-17434

Rendell, M. A., Masters, R. S. W., Farrow, D., & Morris, T. (2010). An implicit basis for the retention benefits of random practice. *Journal of Motor Behavior*, *43*(1), 1–13. https://doi.org/10.1080/00222895.2010.530304

Research and Markets. (2021, May). *The pet training services industry in the United States is expected togGrow at a CAGR of 6% during 2020 to 2026*. https://www.prnewswire.com/news-releases/the-pet-training-services-industry-in-the-united-states-is-expected-to-grow-at-a-cagr-of-6-during-2020-to-2026-301295977.html

Rochais, C., Henry, S., & Hausberger, M. (2017). Spontaneous attention-capture by auditory distractors as predictor of distractibility: A study of domestic horses (Equus caballus). *Scientific Reports*, *7*(15283), 1–8. https://doi.org/10.1038/s41598-017-15654-5

Sanders, G. S., & Baron, R. S. (1975). The motivating effects of distraction on task performance. *Journal of Personality and Social Psychology*, *32*(6), 956–963. https://doi.org/10.1037/0022-3514.32.6.956

Schilder, M. B. H., & van der Borg, J. A. M. (2004). Training dogs with help of the shock collar: Short and long term behavioural effects. *Applied Animal Behaviour Science*, *85*(3), 319–334. https://doi.org/10.1016/j.applanim.2003.10.004

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 179–187. https://doi.org/10.1037/0278-7393.5.2.179

Sheehe, P. R., & Bross, I. D. J. (1961). Latin Squares to balance immediate residual, and other order, effects. *Biometrics*, *17*(3), 405–414. https://doi.org/10.2307/2527834

Smith, P. J. K. (2002). Applying Contextual Interference to snowboarding skills. *Perceptual and Motor Skills*, *95*(3), 999–1005.

Stellato, A., Jajou, S., Dewey, C. E., Widowski, T. M., & Niel, L. (2019). Effect of a standardized four-week desensitization and counter-conditioning training program on pre-existing veterinary fear in companion dogs. *Animals*, *9*(767), 1–18. https://doi.org/10.3390/ani9100767

Todd, Z. (2012, May 2). *How often should I train my dog?* Companion Animal Psychology. https://www.companionanimalpsychology.com/2012/05/how-often-should-i-train-my-dog.html

Utz, R. L. (2014). Walking the dog: The effect of pet ownership on human health and health behaviors. *Social Indicators Research*, *116*(2), 327–339. https://doi.org/10.1007/s11205-013-0299-6

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*(2), 185–211. https://doi.org/10.3758/BF03196276

Xie, Q. (2013). Agree or disagree? A demonstration of an alternative statistic to Cohen’s kappa for measuring the extent and reliability of agreement between observers. *In Proceedings of the Federal Committee on Statistical Methodology Research Conference*, 12.