Odor Span Task in Purpose-Bred Detection Dogs

by

Sarah Marie Krichbaum

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 3, 2019

Keywords: non-match-to-sample, working memory capacity, intervening

Approved by

Jeffrey Katz, Professor of Psychology
Dominic Cheng, Assistant Professor of Psychology
Paul Waggoner, Co-Director of Canine Performance Sciences

Abstract

Working memory is essential for organisms to solve problems related to their survival and to adapt to changes in their environment. Researchers have sought to create a non-human model of working memory that can be used to better understand its predictive value and underlying brain function. Several of these studies have been conducted using the *odor span task* (OST) with rodents, but domestic dogs may be more suitable. Due to the unique similarities between the aging brains of humans and dogs, a model of working memory in dogs could provide useful insight to both veterinary and human medicine. Additionally, the development of an OST in detection dogs is essential to understanding the predictive value of working memory for odors on work success. Six purpose-bred detection dogs were trained on a dog-adapted OST with 24 trials. The OST is a non-match-to-sample task in which the dogs were presented with both a novel (S+) and a previously encountered (S-) odor on each trial. A response to the novel odor was always reinforced. Upon meeting training criterion, the dogs were tested on the OST with an increasing number of odors within a session in order to evaluate working memory capacity (WMC) for odors. The dogs displayed accurate performance (80 percent correct or higher) for set sizes up to 72 odors, which suggested a capacity that is much higher than human WMC. Further analyses focused on the effect of intervening odors (i.e., the number of trials since the S- was last encountered). The dogs demonstrated above chance performance for up to 8 intervening odors, which may be a more accurate representation of dog WMC for odors.

Table of Contents

List of Tables

List of Figures

Odor Span Task in Purpose-Bred Detection Dogs

Executive functions are a group of high-level cognitive processes that are critical for survival and supported by the pre-frontal cortex (Bizon et al., 2012). A number of mechanisms, including working memory, can be used to control the goal-directed behavior that is the main characteristic of executive functioning (Robbins, 1996). Although the definition of working memory in humans is debated, it is widely agreed to be a limited amount of information that is temporarily available for use in current information processing (Cowan, 2017). Working memory allows organisms to behave appropriately in changing environments and to plan for the future (Banich et al., 2004). Measures of working memory have been shown to reliably predict reasoning skills, learning ability, and other aspects of intelligence (Lèpine & Camos, 2005). Additionally, deficits in working memory can be indicative of neurodegenerative disorders such as dementia and Alzheimer's (Studinski et al., 2006).

Human working memory is typically measured as a capacity (i.e., working memory capacity (WMC)) for a number of items. The tasks used to evaluate WMC vary based on research domain, however, they all focus on humans' ability to temporarily store a limited amount of information (Conway et al., 2005; Scharfen et al., 2018). The digit span task was the first to determine that humans have an average working memory capacity of seven plus or minus two items (Miller, 1957). In this task, the participant is presented with a series of digits (one at a time). After the presentation, the participant must repeat back, in order, the digits from the list. The number of digits that they can repeat back before making a mistake is considered their WMC. In the many years since the identification of the "magic number", a number of working memory tasks have supported the claim that this process has a very small capacity (Daneman & Carter,

1980; Kirchner, 1958; Turner & Engle 1989). It is also agreed that items can be chunked in ways that allow humans to remember more than seven items (e.g., a phone number) (Chase & Simon, 1973; Scharfen et al. 2018).

In the following sections, the findings and limitations of previous studies on working memory in non-humans are explored. As well as the advantages of studying WMC for odors in purpose-bred detection dogs. The methods and restrictions of the OST as a measure of WMC is also discussed.

Working Memory in Dogs

Working memory in non-humans can be broadly defined as short-term memory for stimuli within a given testing session (Baddeley, 1995; Honig 1978; Olton & Samuelson, 1976). Although working memory has been thoroughly investigated in many species (e.g., rats, pigeons, and primates), the methods used to operationalize working memory differ across them (Dudchenko et al., 2013). Therefore, a model of working memory in dogs would allow researchers to assess the across species reliability of the OST and provide necessary insight on how working memory is operationalized in the task.

Additionally, recent studies have indicated that the aging dog brain demonstrates the same morphological and cognitive deficits as shown in humans with neurodegenerative disorders (Tapp et al., 2003). Veterinarians have discovered Cognitive Dysfunction Syndrome (CDS) in dogs that increases with age and shares features with human diseases such as dementia and Alzheimer's (Bain et al., 2001). Currently, veterinarians rely on pet owner evaluations to determine this diagnosis (Head et al., 2008). However, a battery of cognitive tests, including one for working memory, could select for dogs that are at risk for or in early stages of CDS. This would improve

the number of dogs that are diagnosed and properly treated while providing information for handling similar human diseases.

Understanding working memory in dogs may have implications for specific groups of working dogs. For instance, detection dogs occupy several roles in our society, from explosive detection to disease diagnosis (Cobb et al., 2015). These roles are highly important in the protection and well-being of our citizens, yet there is nearly a fifty percent dropout rate in training programs (Cobb et al., 2015; Maejima et al., 2007). There is a clear need for improving detection dog selection and training in order to increase the number of dogs successfully placed in service. Because of the complex cognitive processes involved in many detection dog roles, researchers have expressed the need to increase our understanding of cognition in relation to the skills they provide. Given the relationship between working memory and intelligence, understanding its involvement in detection dog work may immensely improve the selection and training of detection dogs, increasing the security of our nation and of our citizens while at home and abroad (Maclean & Hare, 2015).

The current work on dog working memory is dominated by studies that evaluate the temporal duration of working memory (Bensky et al. 2013). In these studies, a dog must locate a stimulus after a delay, requiring retention of some stimulus property across a duration of time. The most common of these procedures is the *variable delay non-match-to-sample task* (vDNMS) (Adams et al., 2000a; Chan et al., 2002; Head et al., 2008; Milgram et al., 1994; Tapp et al., 2003; Salvin et al., 2011; Zanghi et al., 2015). In these procedures, the dog is presented with a single or set of objects. The object(s) are then separated from the dogs' view for a varying delay interval. Following the delay, the dog is required to make a choice regarding the objects and working

memory is recorded as the length of the delay. An incorrect response may indicate that the dog is unable to remember instances after such a delay.

The *variable non-match to position* (vDNMP) procedure, a derivative of the vDNMS, was one of the first memory procedures used in dogs and examined the effects of aging on spatial working memory in both young and old beagles (Adams et al., 2000a). The dog was presented with a tray that held an object (e.g., a cube) on either the left or right side of the tray. The dog received a food reward for displacing the object. Upon receiving the reward, the tray was removed from the dog's view for a varying delay interval. Following the delay, the tray was reintroduced to the dog. This time the original object and an identical object (e.g., another cube) were on the tray (one on the right side of the tray and one on the left). The dog received a food reward for displacing the cube that was in the "novel" position.

The duration of working memory was measured as a function of the delay between the response to the sample and the re-deliverance of the tray. The results indicated that all dogs showed decreased performance with increasing delay. However, a significant age effect was found. Young dogs displayed better working memory than old dogs across delays of 20, 70, and 110 seconds. This finding provided support for the vDNMP as a way to evaluate age-dependent cognitive dysfunction in dogs. Specifically, it provided researchers with a method to separate dogs in to groups based on their memory impairment which could be used as a tool to investigate the relationship between cognitive dysfunction and neuropathology (Adams et al., 2000a).

The vDNMP has since been used to study the predictive value of working memory on other cognitive impairments that are related to aging. Zanghi et al. (2015) used the vDNMP to test whether working memory impairment predicted impairment in

stimulus discrimination and motor learning in old beagles. First, each dog underwent

the normal vDNMP task in order to establish a baseline and separate the dogs into

groups. Each dog was placed in to one of three groups: low memory performance

(LMP), medium memory performance (MMP) and high memory performance (HMP)

group) based on their results on the vDNMP task.

Following the spilt, each dog completed an *object oddity task* to test their ability

to discriminate between visual stimuli. During acquisition of the task, the dog was

presented with a tray that held two objects (e.g., a banana and a block). A response to

one of the objects was always reinforced (e.g., the banana). The testing phase consisted

of two session types; same-distractor and different-distractor. On same-distractor

sessions the dog was presented with a tray that contained the object that was always

reinforced during acquisition and one, two, or three distractor objects. It is important to

note that during this phase the distractors were always identical (e.g., one, two, or three

blue blocks). A response to the object presented during acquisition was reinforced. On

different-distractor sessions the dog was presented with a tray that contained the object

that was always reinforced during acquisition and one, two, or three distractor objects. It

is important to note that during this phase the distractors were never identical (e.g., a red

triangle, a blue block, and a green sphere). A response to the object presented during

acquisition was reinforced. The dogs were evaluated on their ability to choose the item

that was always reinforced despite distractions.

Next the dogs were retested on vDNMP to ensure that the time that elapsed since

the first test did not affect their performance on the retest. The results showed no

significant differences between the two vDNMP tests and confirmed that vDNMP was a

reliable measure of working memory. Lastly, the dogs completed a *motor learning task*.

Each dog was placed in a chamber that contained a hole where their paw could extend out. On the first trial, a tray that contained a food reward was placed directly in front of the hole. If the dog used a paw to pull in the tray the dog was reinforced. On each day thereafter, the tray was placed farther away from the chamber hole. The dogs were evaluated on the maximum distance that they learned to pull in the tray which was reflective of motor learning.

The results from the *object oddity task* revealed a significant main effect for the number of distractor objects during both same-distractor and different-distractor conditions. However, there was not a significant effect for group. In other words, as the number of distractors increased the dogs' response accuracy decreased at a similar rate for each group. The results from the *motor learning task* indicated that the maximum distance learned to contact the reward tray was not significantly different between groups. In conclusion, results from this study showed that working memory performance does not reliably predict some related processes (e.g., discrimination and motor learning). Therefore, these tasks in part assess different functions (Zanghi et al., 2015).

Another procedure used to measure working memory in dogs is the *visible displacement task* (Fiset et al., 2003). This task specifically measures the memory component that is involved in object permanence (i.e., knowing that an object still exists when it is no longer viewable in the environment). In this study, the dog watched an experimenter hide an object behind one of four boxes. Then the experimenter placed an opaque screen between the dog and the boxes for a varying delay interval. Following the delay, the screen was removed, and the dog was required to locate the hidden object. The results indicated that dogs were able to locate the hidden object at above chance

levels (25%) for delays up to 240 seconds. The authors attributed this to the dogs working memory, however, there are important limitations to consider. First, because the dogs were able to locate the hidden objects following the longest delay, the maximum duration of the dogs working memory is still unknown. It also raises the idea that the dogs could have been relying on a more long-term memory system. Second, the dogs could have used a non-mnemonic strategy, such as orienting their body toward the correct location in order to solve the task. Third, the dogs could have used scent detection, or their ability to smell the reinforcement behind the correct location in order to solve the task (Fiset et al., 2003).

Though informative, these studies may not be ecologically valid and lacked important features of the tasks performed by detection dogs First, in these tasks dog working memory was measured as a function of delay rather than storage capacity. In other words, dog working memory was determined by how long (i.e., the duration) the dog could remember the location of a hidden stimulus rather than how many (i.e., the capacity) stimuli the dog could remember in a short period of time. Detection dogs are required to remember many odors when introduced to complex environments; therefore, proficient WMC is a necessity. A measure of capacity is also important for comparing the findings to working memory in humans. Additionally, there is evidence that dogs perform better on long-term memory tasks that use odor stimuli instead of visual stimuli, therefore, it seems more appropriate to use dogs' dominant sensory modality (Hall, 2015).

Odor Span Task

An *odor span task* has the potential to fulfill some of these requirements. The odor span task has been used to examine rats' working memory capacity for odors

(April et al., 2013). In April and colleagues, on trial one, the rat was placed in the center of an arena that contained eighteen holes. One hole contained a cup with an odorized lid (S+) and the other seventeen holes contained empty cups. Responding to the odorized lid (flipping the lid) was reinforced. On trial two and on, two holes contained cups with odorized lids (a novel odor: S+ and an odor that the rat had been presented with on a previous trial: S-). Responding to the novel odor (flipping the lid) was reinforced. During Experiment 1, the rats completed 24 trials per session, meaning that they were required to hold 24 odors in their working memory. The results from Experiment 1 showed that the rats responded significantly above chance for up to 24 odors. In Experiment 2, the researchers sought to find rats' working memory capacity for odors. The rats completed three session types (36 trials, 48 trials, and 72 trials) over two weeks (1 or 2 sessions per type). The results indicated that all ten of the individual rats had a mean percent correct that was significantly above chance for up to 72 odors to remember. When examining accuracy across session there was a significant decrease across trial block (blocks of 12 trials). However, the rats maintained above chance accuracy throughout.

This large capacity strongly contradicted the notion that the OST could be used as a model of the limited human working memory capacity (April et al., 2013). The authors also suggested that the rats were using relative familiarity (or choosing the odor that was the least familiar) rather than working memory to solve the task. The importance of using the OST to evaluate working memory in providing insight on these issues is two-fold. First, across species comparison of the OST can be determined between dogs and rats. Second, working memory will not only be evaluated as success

across the length of the session but also across the number of intervening odors, which may be a more accurate measure of non-human WMC (Kirchner, 1958).

The methods of the current study were based on those explained above (April et al., 2013). A dog-adapted OST was created in order to assess working memory in detection dogs. Dogs underwent OST training with session lengths of 24 trials. It was hypothesized that dogs would successfully learn the OST and display high accuracy for up to 24 odors. Next, the session lengths were incrementally increased in order to evaluate the dogs' WMC. It was hypothesized that dogs would display high accuracy for up to 72 odors, similar to what was found in rats (April et al., 2013).

Method

Subjects

Six purpose-bred detection dogs from the Auburn University Canine Performance Sciences (CPS) program were used in this study. The dogs varied in age ($M = 3.66$ years), gender (3 = Females, 3 = Males), and status in the program. Ethical approval for the study was obtained from the Auburn University Institutional Animal Care and Use Committee. Each dog was previously trained on explosive detection and participated in odor detection studies. However, only one dog was previously trained with any of the odors used in the current study. In that case, the dog was trained to never respond to five of the current odors (i.e., almond, apple, cinnamon, onion, and tobacco) but that did not impact the performance of the dog after a few pre-training sessions. Experimental sessions occurred four times a week.

Apparatus

Dog Adapted Arena. All training and testing sessions took place in a 720 x 720 - inch building within a 179 x 163 -inch confined space at the Canine Performance

Science center. Eight, 7 1/2 x 7.5 1/2 x 7.5 ½ -inch cinder blocks, with the open end up, were placed on 11 x 11 x 7-inch wooden boxes and arranged in a square formation. The cinderblocks were placed 27 inches apart and spaced with 48 x ¾ x 12-inch plywood boards. An additional 43 x ½ x 23-inch plywood board, placed between cinderblocks one and eight, ensured that the dogs' sampled systematically during OST Pre-Training (see Figure 1). The holes in each cinderblock were 5 x 5-inches so that the stimulus cans fit securely in the holes. A Vixia HR70 camera was used to record all training and testing sessions.

Figure 1. The dog-adapted arena. Dogs' entered through the opening at the top of the figure, then systematically sampled the wheel (locations 1-8) in a counterclockwise fashion.

Stimuli

The stimuli were odorized cotton rounds that rested in tins, 2 ½ inches in diameter. The tins were perforated with nine 3/32 -inch holes for odor release. The cotton rounds were odorized by storing them in airtight containers with the odors for a minimum of one week. In training, stimuli were selected from a pool of 36 odors. In testing, 36 new odors were added, for a total of 72 odors. All of the odors were ordered from Great

American Spice Company Co., with the exception of tobacco (see Table 1). Upon

stimulus presentation, the tins were placed in larger cans, with openings 3 ½ inches in

diameter, so that they fit securely in the cinderblocks (see Figure 2).

Table 1

Stimuli for training (36 odors, top panel) and testing (36 odors, bottom panel)

Training					
Allspice	Almond	Anise	Apple	Asparagus	Basil
Bay	Caraway	Carob	Celery	Chili	Cinnamon
Coriander	Cumin	Dill	Fennel	Fenugreek	Garlic
Ginger	Mustard	Nutmeg	Onion	Oregano	Paprika
Parsley	Red Pepper	Rosemary	Sage	Savory	Sesame
Sumac	Thyme	Tobacco	Tumeric	Vanilla	Watermelon

Testing					
Amaretto	Apricot	Banana	B. Walnut	Blackberry	Blueberry
Brandy	Butter	Butter Rum	Butterscotch	Caramel	Champagne
Cherry	Chocolate	Clove	Coffee	C. Candy	Eggnog
E. Toffee	Grape	Honey	Lemon	Lime	Maple
Marjarom	Marshmallow	Peach	P. Butter	Pecan	Pineapple
Pistachio	Pumpkin	Raspberry	Rootbeer	Strawberry	Tangerine

Figure 2. A perforated tin inside a cinderblock for stimulus presentation.

Procedure

Phase 1: Acclimation to the Arena. The first few sessions trained the dogs to respond to novel odors. These sessions consisted of 24 trials, and on each trial, a single novel odor and two empty cans were randomly placed in the arena. A nose poke within 5 centimeters of the cinderblock in combination with sitting in front of the position that contained the novel odor for 1.5 seconds was reinforced. The dog's response was marked by a click that cued the dog to exit the arena and receive the reinforcement (Chuck-It ball). Once the dog was successfully sitting in front of novel odors an incrementing non-match-to-sample procedure was implemented.

Phase 2: OST Pre-Training. During this phase, the dog always completed 24 trials in a session (collected on a single day). The session could occur in three blocks of 8 trials, two blocks of 12 trials, or one block of 24 trials with ten minutes between blocks. The blocking was determined by the dog's performance on that day, as to whether a break was needed. For example, if a correction procedure (described below) was used multiple times, the trials occurred in smaller blocks to prevent fatigue. The purpose of the correction procedures was to address any position biases that the dogs had (e.g., previous training to not respond to certain locations). By the end of Phase 2, all of the dogs were completing 24 trials per session without any breaks.

On the first trial of each session, a single odor was placed in a random location in the arena. A handler released the dog in to the arena with the previously trained command "search". When the dog sat in front of the odor, the response was marked with a click and the dog exited the arena to receive the reinforcement. On every subsequent trial, the dog was presented with a two-choice discrimination between a previously encountered odor from that session (S-), selected by the experimenter, and a

novel odor (S+). The S+ odors were randomly selected from the pool of stimuli with the constraint that the odor did not previously occur in a session. Both odors (S+ and S-) were placed in random locations on the wheel. The dog received reinforcement for responding to the novel odor.

Two types of correction procedures were used in this phase. First, a time-out procedure was implemented so that if the dog sat in front of the previously encountered odor (S-), the handler called the dog to exit the arena with a previously trained command "come", held the dog for 10 seconds, then released the dog back in to the arena. The dog remained in the arena until responding to the novel odor. Second, a wait-out (i.e., autocorrection) procedure was implemented so that if the dog sat in front of the previously encountered odor (S-) the dog remained in the arena until responding to the novel odor. The type of the correction procedure used depended on individual dog need. Once a performance criterion of a minimum of ten sessions with at least 75 percent correct on a session of 24 trials was obtained, dogs advanced to Phase 3.

Phase 3: OST Training. All of the OST Training sessions consisted of 24 trials and were identical to OST Pre-Training except that the previously encountered odor (S-) was randomly selected with replacement. During each session, 6 of the trials were considered odor control trials. On these trials, the previously encountered odor (S-) was in a new tin and a new can. These control trials ensured that when a dog encountered this odor it was free of the dogs' own odorants (i.e., scent marking). That is, these trials controlled for the possibility of the dogs rejecting an odor because they could smell themselves. Intermittent, non-randomized maintenance sessions were conducted if the dogs presented biases (e.g., avoiding a specific location). Each dog was required to meet an OST Training criterion of a minimum of ten sessions with at least 84 percent

correct for two consecutive sessions (with no correction procedure) before advancing to the next phase. If there was more than a two-day break in testing following criterion, the dogs were required to meet one additional day of at least 84 percent correct on a session of 24 trials (deemed maintenance sessions). Dogs' WMC was measured by expanding the set from 24 up to 36, 48 and 72 odors.

Phase 4: Exposure. Upon completion of OST Training in Phase 3, two of the dogs completed an exposure phase. This phase was added prior to the Working Memory Capacity Test in Phase 5 in order to determine if exposure to the entire odor set had an effect on performance. It was possible that rapidly increasing the number of odors in Phase 5, might have a deleterious influence on performance due to the increase in novel odors. If so, dogs in Phase 4 would perform better than dogs not exposed to the odors. Hence, to determine if novelty was an important variable, the other four dogs did not undergo the exposure phase before moving to Phase 5. Each session was 24 trials. This phase consisted of nine Phase 3 training sessions and its purpose was to expose the dog to all of the 72 odors that would be presented during Phase 5. Therefore, the 72 stimuli were balanced for number of presentations across the nine exposure sessions. If there was more than a two-day break in testing following exposure, the dogs were required to meet one additional day of performance criterion of 84 percent correct on a session of 24 trials (deemed maintenance sessions). Upon meeting this criterion, the dogs advanced to Phase 5.

Phase 5: Working Memory Capacity Test. All of the dogs completed the Working Memory Capacity Test. The sessions occurred the same as the OST Testing sessions except that the number of trials varied between 36, 48, and 72. The number of trials was equivalent to the number of odors that were tested, such that dogs were

required to remember up to 72 odors. The odor stimuli were selected and set in the arena in the same way with the addition of 36 new odors (see Table 1). Each dog completed two sessions of each stimulus condition. During the first week of testing, the sessions were conducted in ascending order (i.e., 36, 48, then 72 odors). During the second week, that order was repeated. Additionally, correction procedures were not implemented during this phase.

Data Analysis. Percent correct was calculated as the number of correct responses divided by the number of trials in each session. A correct response was defined as sitting in front of the novel odor (S+) without a response to the previously encountered odor (S-) or an empty can. The main variable of interest was percent correct for number of odors in a session. Additional analyses focused on the number of intervening odors (i.e., the number of trials since the S- was last encountered) and the number of spaces between the S+ and S- (i.e., the number of empty stimulus locations between the S+ and S-) on percent correct.

Data sheets were created for OST Pre-Training, OST Training, Exposure, and the Working Memory Capacity Test (for an example see Appendix I) for live scoring. The data sheets provided trial-by-trial information for the specific odor and stimulus location. Additionally, twenty percent of the testing videos were scored by a second independent observer. This observer agreed with the live scorer on 100% of these trials, as the response was well defined in each dog.

Results

Training. The number of total sessions in each phase varied due to the number of sessions it took each dog to meet criterion in the OST Pre-Training (M = 17.17, SEM = 2.10) and OST Training (M =18.83, SEM = 3.04) phases (see Table 2). All of the dogs

acquired the task and met the OST Training performance criterion of two consecutive sessions of 84 percent correct or above. Accuracy on the second to last (M = 88.89, SEM = 2.56) and last (M = 86.81, SEM = 1.99) sessions of OST training were not statistically different from the criterion value of 84 percent correct, as determined by one-sample *t*-tests $t(5) = 1.91$, $p = .12$ and $t(5) = 1.41$, $p = .22$, respectively.

Table 2

Number of sessions that each dog completed by phase.

Subject	OST Pre-Training	OST-Training	Maintenance	Exposure	WMC Test	Total
Beaufort	13	32	16	9	6	76
Jolly	16	16	19	9	6	72
Kylee	16	20	2	-	6	44
Nessie	20	10	8	-	6	44
Quail	26	15	6	-	6	53
Vera	12	20	1	-	6	39

Working Memory Capacity Test. Figure 3 shows the average and individual dog performance across the three stimulus set sizes. Overall, the dogs displayed comparable accuracy for the 36, 48, and 72 odor set sizes as confirmed by a two-way repeated measures ANOVA of set size (36, 48, 72) and test order (first, second) which revealed no main effect of test order $F(1, 5) = .07$, $p = .80$, or set size $F(2, 10) = 1.97$, $p = 1.19$, and no significant interaction $F(2, 10) = 2.06$, $p = .18$. Given that performance on each set size was not influenced by test order the data within each set size was averaged for the remaining analyses. The dogs exhibited above chance accuracy (50%) on set sizes of 36

stimuli, as determined by one-sample *t*-tests, $t(5) = 25.08$, $p < .001$, 48 stimuli, $t(5) = 37.35$, $p < .001$, and 72 stimuli, $t(5) = 21.23$, $p < .001$. There was no difference in accuracy between the dogs (Beaufort and Jolly) that experienced the entire odor set prior to testing in Phase 4 and the dogs (Kylee, Nessie Quail, and Vera) that did not experience the entire set prior to testing. Therefore, all of the dogs were included in the analyses.
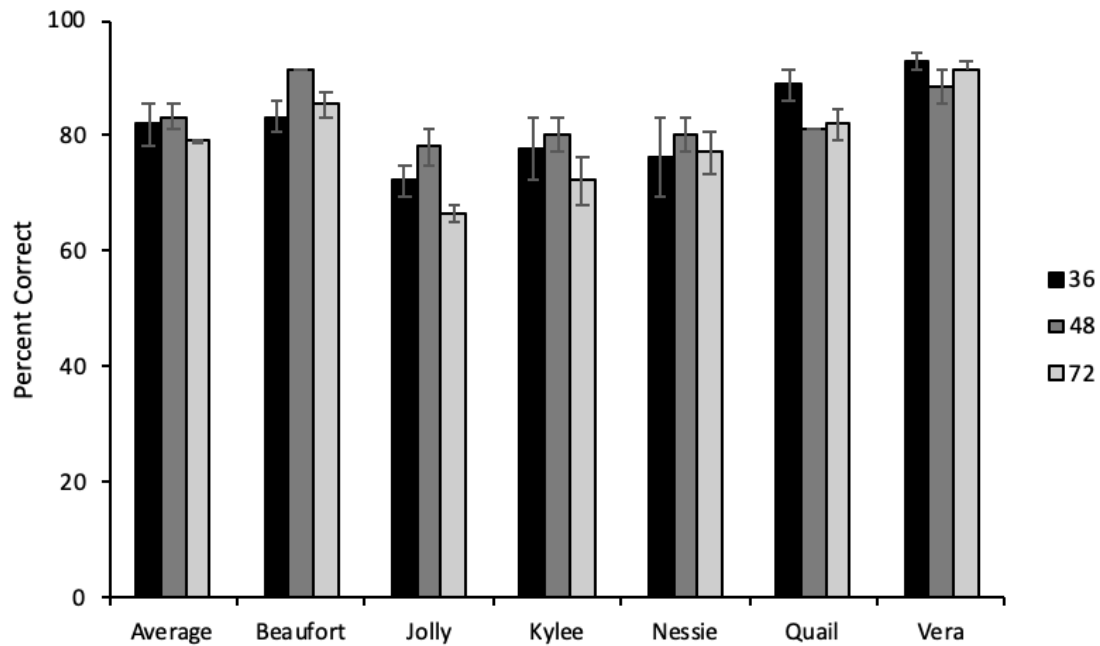

Figure 3. Average and individual dog performance shown as overall percent correct for each set size (36 stimuli: black bars; 48 stimuli: gray bars; 72 stimuli: light gray bars), error bars represent SEM.

Across session performance for each set size across 6-trial blocks can be seen in Figure 4. The dogs displayed a decrease in accuracy across trial blocks for set sizes of 36, 48 and 72, which was similar for each set size. A series of ANOVA's supported these findings. A two-way repeated measures ANOVA for trial block (1-6) and set size (36, 48, 72) revealed a main effect of trial block $F(5, 25) = 9.69$, $p < .001$, no main effect of set size $F(2,10) = .807$, $p = .47$, and no interaction $F(10, 50) = .41$, p = .94. Similarly, a two-

way repeated measures ANOVA for trial block (1-8) and set size (48, 72) revealed a main effect for trial block $F(7, 35) = 5.05$, $p < .001$, no main effect for set size $F(1,5) = .02$, $p = .90$, and no interaction $F(7,35) = .35$, $p = .93$. Additionally, a one-way ANOVA for trial block (1-12), for the set size of 72, also revealed a main effect of trial block $F(11, 55) = 4.51$, $p < .001$. These findings suggest that the decrease in accuracy across trial block was the same for all session lengths. Importantly, the dogs maintained above chance accuracy for all trial blocks for each set size, as confirmed by Bonferroni-corrected ($\alpha = .05/26 = .002$) one-sample t-tests, ps $< .001$.
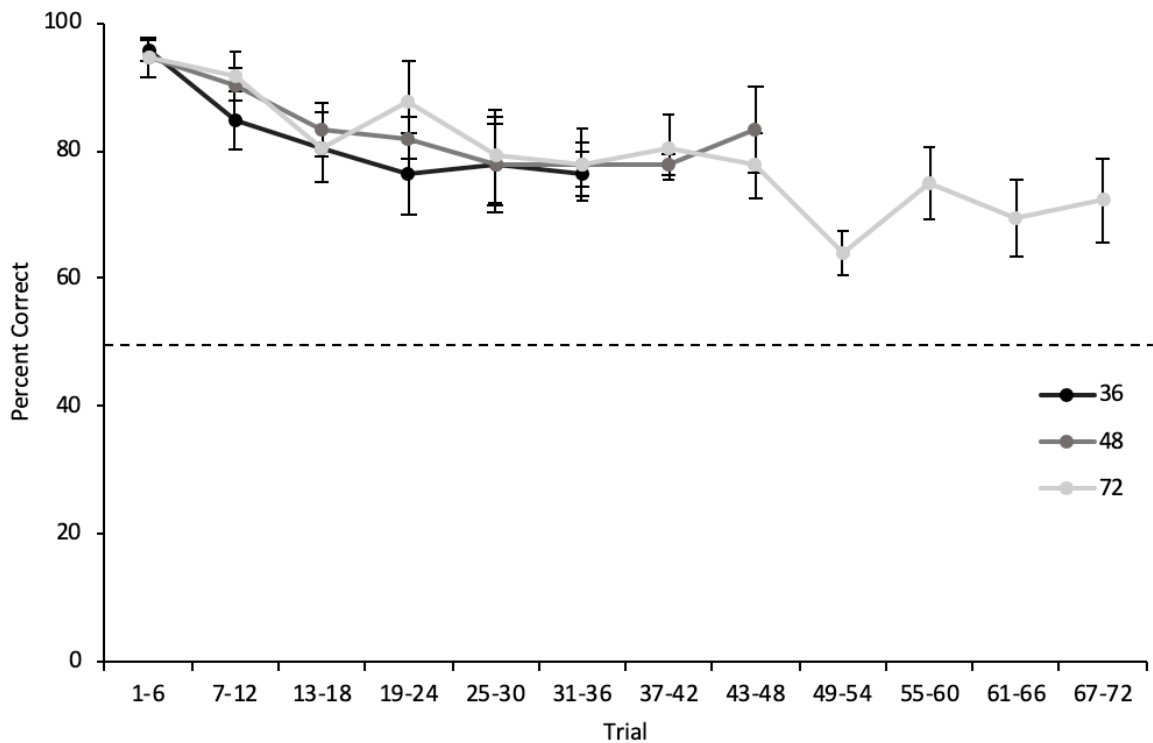

Figure 4. Average percent correct across trial (blocks of 6 trials) for each set size (36 stimuli: black circles; 48 stimuli: dark gray circles; 72 stimuli: light gray circles), error bars represent SEM.

Odor Controls. Baseline and odor control trials were compared throughout testing in order to insure that there was no effect of scent marking which was supported by a

18

paired samples *t*-test that revealed no significant difference between them $t(5) = -1.02$, *p* = .35. The dogs scored significantly above chance on baseline (M = 80.67, SEM = 1.84) and odor control trials (M = 84.26, SEM = 3.87) as confirmed by one-sample *t*-tests (p < .001).
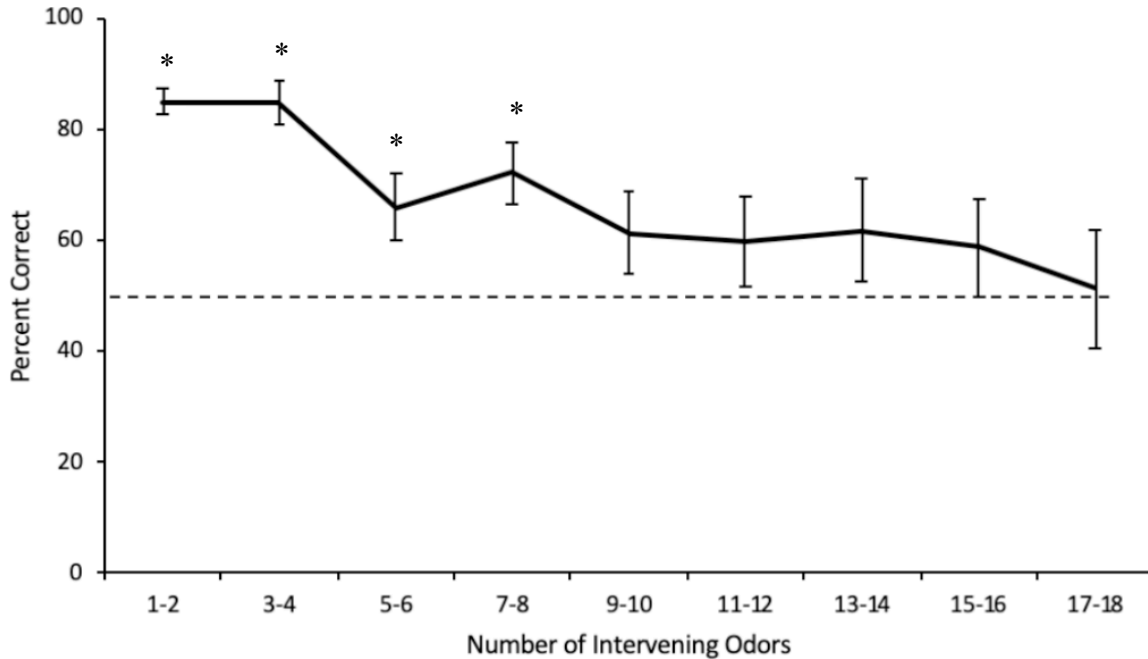

Figure 5. Average percent correct across number of intervening trials (blocks of 2) across all test sessions (* indicates significant binomial test, *p* < .05).

Intervening Odors. Figure 5 depicts performance as a function of the number of intervening odors during testing sessions. Trials in which the dog did not encounter the S- (as determined by coded number of visits to the S-) were excluded from this analysis because exposure to the S- odor could not be verified. The dogs showed a decrease in accuracy across number of intervening odors as supported by a one-way repeated measures ANOVA that demonstrated a main effect of number of intervening odors block $F(8, 96) = 2.11$, $p = .04$. Although there was a decrease in accuracy across number of intervening odors, the dogs displayed significantly above chance accuracy for up to 8

intervening trials as confirmed by Bonferroni-corrected (α = .05/9 = .006) one-sample *t*-tests, *p*s < .001.

      Spacing. The effect of number of spaces between the S+ and S- on accuracy during testing is shown in Figure 6. There was no difference in accuracy across number of spaces as disclosed by a one-way repeated measures ANOVA on accuracy that revealed no main effect of number of spaces, $F(3, 51)$ = .40, p = .75.
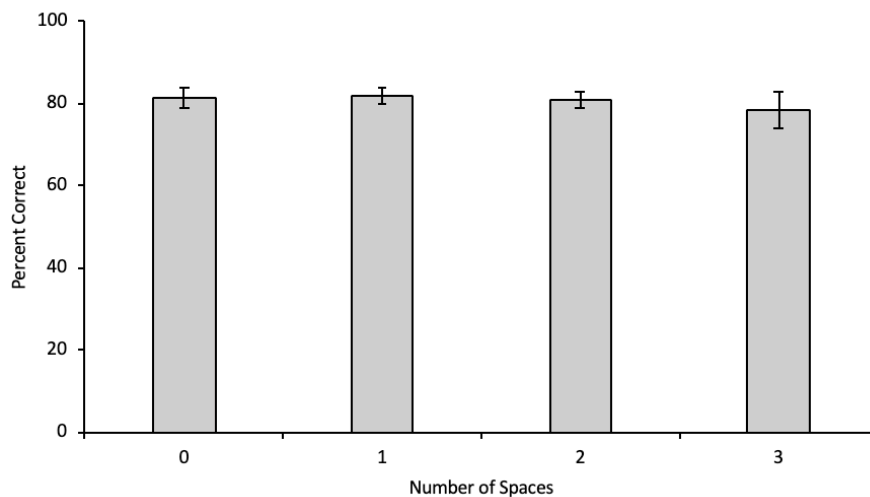

Figure 6. Average percent correct across number of spaces between the S+ and S- for all testing sessions.

Discussion

      All dogs learned the OST and performed at high levels (>84%) on the last two OST Training sessions before moving on to testing with increased set sizes. These results supported the hypothesis that a group of purpose-bred detection dogs could learn the two-choice non-match-to-sample-task. This high level of performance continued throughout testing and was not affected by test order (first or second), the size of the stimulus set (36, 48, 72), or exposure to the set of odors prior to testing. These findings supported the hypothesis that dogs would exhibit high accuracy for up to 72 odors. Additionally, dogs displayed above chance accuracy for up to 8 intervening odors during testing. Other

variables, including spacing of the S+ and S- and odor controls (ruling out scent marking), did not affect accuracy during testing.

As expected, dogs performed similar to rats when across session accuracy was examined across each set size (Figure 7). Dogs showed a decrease in accuracy across trial block for set sizes of 36, 48, and 72. In other words, dogs' performance on the OST decreased as a function of session length but overall accuracy was the same for each stimulus set (cf., April et al., 2013). Similar to what was found in rats, the slopes of all three functions are shallow and the dogs remained well above chance for up to 72 trials. It is possible that overall accuracy was comparatively high in dogs, like rats, because they have unlimited working memory for odors (or at least up to 72). Additional species should be tested on versions of the span task that implement different stimulus modalities (e.g., a visual span task in pigeons) in order to determine if these results are specific to odor memory.
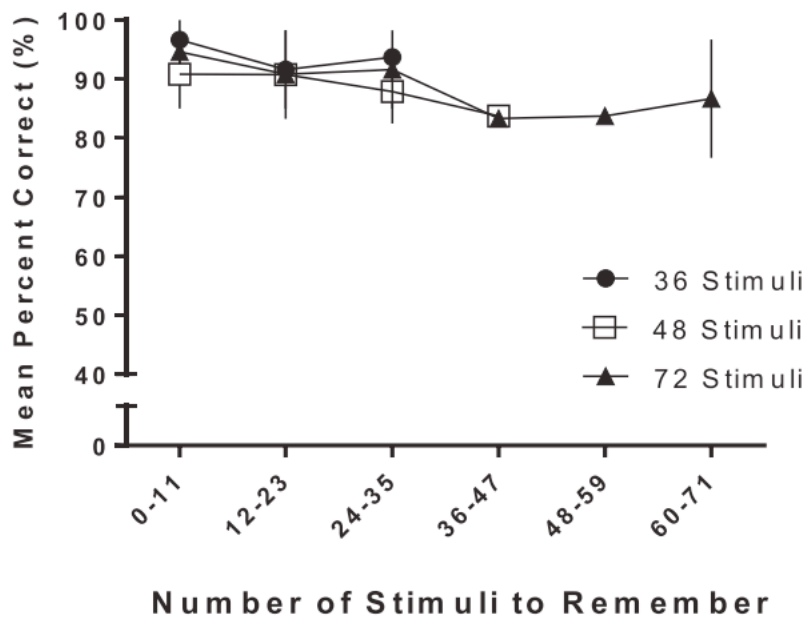

Figure 7. Average percent correct across trial (blocks of 6 trials) for each set size (36 stimuli: black circles; 48 stimuli: white squares; 72 stimuli: black triangles), error bars represent SEM. After April et al., 2013.

When examining memory this way, it appears that dogs have the ability to hold a large number of odors in working memory. However, it is possible that the OST does not measure working memory as suggested by April et al. 2013. That is, the dogs could have responded based on the process known as relative familiarity. Familiarity works parallel to working memory but only encodes if a stimulus has been encountered previously (e.g., Basille & Hampton, 2013; Pañoz-Brown et al., 2016; Yonelinas, 2002;). For example, the dogs could have avoided the comparison odor due to a sense of "oldness" (April et al., 2013). Research suggests that animals can perform well on memory tests by relying solely on familiarity and that it may have an infinite capacity unlike its working memory counterpart (Basille & Hampton, 2013). Future work should seek to tease apart these constructs in an effort to better understand the mechanism of working memory (c.f., & Hampton, 2013; Brady & Hampton, 2018).

Another possibility is that WMC could be assessed as the number of intervening odors. In other words, the number of discriminations since the S- was last encountered. This analysis is similar to the n-back task that measures WMC in humans (Kirchner, 1958). In a typical n-back task a participant is presented with a list of items. Following presentation of the list, the participant is asked if one of the items was presented n-back in the list. For example, during a two-back trial a participant may be shown a series of pictures then asked if a specific picture was presented two times ago or second to last (Redick & Lindsey, 2013). WMC is probed by increasing the n-back. Analyzing intervening odors in the OST may be similar to assessing n-back as dogs are required to remember if the S- odor was encountered across many trials. Interestingly, dogs displayed above chance accuracies for up to 8 intervening odors and suggests that the OST may be a measure of WMC for odors in dogs. It is important to note that the number

of observations in each block of intervening odors was not counterbalanced. Specifically, the as the number of intervening odors increased the number of observations decreased, as represented by the error bars in Figure 5. Future work should implement a counterbalancing technique in order to create an equal number of observations across the number of intervening trials (cf. Wright, Katz & Ma, 2012).

Detection dogs' ability to quickly learn and locate odors in changing environments made them a good candidate for this task. However, due to their previous training it was important to control for their ability to use extraneous odor cues. During testing, the dogs showed no difference in accuracy on odor control and baseline trials meaning they did not avoid the S- odor due to possible scent marking of the stimuli. Another concern was that due to the close proximity of the cinderblocks the dogs could smell both the S+ and S- odors upon entering the arena. Therefore, it was hypothesized that the closer the odors were spaced in the arena the more difficult it would be for the dog to pinpoint odor source and that this would be represented by lower accuracy on those trials compared to trials in which the odors were spaced farther apart. Upon analysis, there was no effect of stimuli spacing suggesting that the dogs learned to pinpoint odor source regardless of if they could smell the odors upon entering the arena.

The results of this study provided insight on the working memory abilities of detection dogs. Future studies should explore the connection between performance on the OST and job placement as it is possible that performance on the OST early in development could have predictive value on detection success, as recent studies have indicated the significance of using cognitive measures, including working memory, on similar assessments (Maclean & Hare, 2015). Though, many detection dogs rely on long-term memory for previously trained target, some, such as human scent line-up dogs, are

required to make many odor discriminations within a short period of time (Jerzierski et al., 2010; Settle et al., 1994). In these cases, dogs are trained on a match-to-sample procedure in which the dog is required to sample an object that was associated with a crime and then select the suspect amongst a number of individuals (Schoon, 1996). Similar to the OST, the match-to-sample procedure should be considered for future studies as it may have more direct applications to detection dog work.

The OST as a model of working memory in dogs may hold value in both veterinary and human medicine. As previously discussed, working memory deficits are common to both Canine Cognitive Dysfunction and human neurogenerative disorders (Zanghi et al., 2015). Therefore, dozens of studies have been devoted to finding a model of working memory in dogs in which performance is negatively affected by age (Adams et al., 2000a; Chan et al., 2002; Head et al., 2008; Milgram et al., 1994; Tapp et al., 2003; Salvin et al., 2011; Zanghi et al., 2015). Future work should examine dogs' performance on the OST at different ages as it may be a valid measure of cognitive decline that could be used to detect symptoms of CDS and similar human diseases.

Conclusions

The results of the current study offer useful information regarding the ability of detection dogs to remember odors over a short period of time. First, the dogs' displayed high accuracy across 72 trials which is indicative of their ability to learn non-match-to-sample with odors but also to make 72 discriminations in a single session. Second, when the testing sessions were evaluated in terms of intervening odors, dogs succeeded in making the correct choice following 8 intervening odors since the S- was encountered which may be indicative of their WMC for odors. Although it is necessary to dissociate the processes of familiarity and working memory on these tasks, the OST may have

important value in the training and selection of detection dogs as well as to inform both veterinary and human medicine.

References

Adams, B., Chan, A., Callahan, H., & Milgram, N. W. (2000). The canine as a model of human cognitive aging: recent developments. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*.

April, L. B., Bruce, K., & Galizio, M. (2013). The magic number 70 (plus or minus 20): variables determining performance in the rodent odor span task. *Learning and Motivation*, *44*(3), 143-158.

Baddeley, A. (1995). Working memory. In *The Cognitive Neurosciences,* ed. M. Gazzaniga, 755-764. Cambridge, MA: MIT Press.

Bain, M. J., Hart, B. L., Cliff, K. D., & Ruehl, W. W. (2001). Predicting behavioral changes associated with age-related cognitive impairment in dogs. *Journal of the American Veterinary Medical Association*, *218*(11), 1792-1795.

Banich, M. T. (2004). *Cognitive Neuroscience and Neuropsychology*. Houghton Mifflin College Division.

Basile, B. M., & Hampton, R. R. (2013). Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition*, *126*(3), 391-396.

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, pp. 209-406). Academic Press.

Bizon, J. L., Foster, T. C., Alexander, G. E., & Glisky, E. L. (2012). Characterizing cognitive aging of working memory and executive function in animal models. *Frontiers in Aging Neuroscience*, *4*, 19.

Brady, R. J., & Hampton, R. R. (2018). Nonverbal working memory for novel images in rhesus monkeys. *Current Biology*, *28*(24), 3903-3910.

Chan, A. D., Nippak, P., Murphey, H., Ikeda-Douglas, C. J., Muggenburg, B., Head, E.,
Cotman, C. W., & Milgram, N. W. (2002). Visuospatial impairments in aged
canines (Canis familiaris): the role of cognitive-behavioral flexibility. *Behavioral
Neuroscience*, *116*(3), 443.

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.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R.
W. (2005). Working memory span tasks: A methodological review and user's
guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.

Cowan, N. (2017). The many faces of working memory and short-term
storage. *Psychonomic Bulletin & Review*, *24*(4), 1158-1170.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and
reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Dudchenko, P. A., Talpos, J., Young, J., & Baxter, M. G. (2013). Animal models of
working memory: a review of tasks that might be used in screening drug
treatments for the memory impairments found in schizophrenia. *Neuroscience &
Biobehavioral Reviews*, *37*(9), 2111-2124.

Fiset, S., Beaulieu, C., & Landry, F. (2003). Duration of dogs'(Canis familiaris) working
memory in search for disappearing objects. *Animal Cognition*, *6*(1), 1-10.

Hall, N. J., Glenn, K., Smith, D. W., & Wynne, C. D. (2015). Performance of Pugs,
German Shepherds, and Greyhounds (Canis lupus familiaris) on an odor-
discrimination task. *Journal of Comparative Psychology*, *129*(3), 237.

Head, E., Rofina, J., & Zicker, S. (2008). Oxidative stress, aging, and central nervous system disease in the canine model of human brain aging. *Veterinary Clinics: Small Animal Practice*, *38*(1), 167-178.

Honig, W. K. (1978). Studies of working memory in the pigeon. In *Cognitive Processes in Animal Behavior* (pp. 211-248). Routledge.

Jezierski, T., Górecka-Bruzda, A., Walczak, M., Swiergiel, A. H., Chruszczewski, M. H., & Pearson, B. L. (2010). Operant conditioning of dogs (Canis familiaris) for identification of humans using scent lineup. *Animal Science Papers and Reports*, *28*(1), 81-93.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352.

lle Lépine, R., Parrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, *12*(1), 165-170.

MacLean, E. L., & Hare, B. (2015). Dogs hijack the human bonding pathway. *Science, 348*(6232), 280-281.

Maejima, M., Inoue-Murayama, M., Tonosaki, K., Matsuura, N., Kato, S., Saito, Y., Weiss, A., Murayama, Y., & Ito, S. I. (2007). Traits and genotypes may predict the successful training of drug detection dogs. *Applied Animal Behaviour Science*, *107*(3), 287-298.

Milgram, N. W., Head, E., Weiner, E., & Thomas, E. (1994). Cognitive functions and aging in the dog: acquisition of nonspatial visual tasks. *Behavioral Neuroscience*, *108*(1), 57.

Miller, G. A. (1957). The magical number seven, plus or minus two: Some limits on our

     capacity for processing information. *Psychological Review, 101*(2), 343-352

Olton, D. S., & Samuelson, R. J. (1976). Remembrance of places passed: spatial memory

     in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(2),

     97.

Pañoz-Brown, D., Corbin, H., Dalecki, S., Gentry, M., Brotheridge, S., Sluka, C., Wu, J.,

     67 & Crystal, J. (2016). Rats remember items in context using episodic memory.

     Current Biology, 26(20), 2821-2826.

Redick, T. S., & Lindsey, D. R. (2013). Complex span and n-back measures of

     working memory: a meta-analysis. *Psychonomic bulletin & review*, *20*(6), 1102-

     1113.

Robbins, T. W. (1996). Dissociating executive functions of the prefrontal cortex. *Phil.*

     Trans. R. Soc. Lond. B, *351*(1346), 1463-1471.

Salvin, H. E., McGreevy, P. D., Sachdev, P. S., & Valenzuela, M. J. (2011). Growing old

     gracefully—Behavioral changes associated with "successful aging" in the dog,

     Canis familiaris. *Journal of Veterinary Behavior: Clinical Applications and*

     Research, *6*(6), 313-320.

Scharfen, J., Jansen, K., & Holling, H. (2018). Retest effects in working memory capacity

     tests: A meta-analysis. *Psychonomic Bulletin & Review*, 1-25.

Schoon, G. A. A. (1996). Scent identification lineups by dogs (Canis familiaris):

     experimental design and forensic application. *Applied Animal Behaviour*

     Science, *49*(3), 257-267.

Settle, R. H., Sommerville, B. A., McCormick, J., & Broom, D. M. (1994). Human scent

     matching using specially trained dogs. *Animal Behaviour*, *48*(6), 1443-1448.

Studzinski, C. M., Christie, L. A., Araujo, J. A., Burnham, W. M., Head, E., Cotman, C. W., & Milgram, N. W. (2006). Visuospatial function in the beagle dog: an early marker of cognitive decline in a model of human aging and dementia. *Neurobiology of Learning and Memory*, *86*(2), 197-204.

Tapp, P. D., Siwak, C. T., Estrada, J., Head, E., Muggenburg, B. A., Cotman, C. W., & Milgram, N. W. (2003). Size and reversal learning in the beagle dog as a measure of executive function and inhibitory control in aging. *Learning & Memory*, *10*(1), 64-73.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*(2), 127-154.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441-517.

Wright, A. A., Katz, J. S., & Ma, W. J. (2012). How to be proactive about interference: Lessons from animal memory. *Psychological science*, *23*(5), 453-458.

Zanghi, B. M., Araujo, J., & Milgram, N. W. (2015). Cognitive domains in the dog: independence of working memory from object learning, selective attention, and motor learning. *Animal cognition*, *18*(3), 789-800.

Subject:						
Experimenter:						
Session:						
Date:						
Trial	Stimulus	S+ Location	S-	S- Location	Response	Visits to S-
1	Vanilla	4				
2	Rosemary	5	1	8		
3	Caraway	6	2	4		
4	Thyme	3	3	5		
5	Allspice	6	2	7		
6	Anise	6	4	4		
7	Parsley	4	5	6		
8	Tobacco	3	7	1		
9	Oregano	5	7	2		
10	Bay	3	5	5		
11	Garlic	8	1	6		
12	Carob	6	7	2		
13	Paprika	3	4	5		
14	Savory	2	13	3		
15	Tumeric	4	12	2		
16	Mustard	3	4	5		
17	Watermelon	8	10	7		
18	Ginger	4	17	2		
19	Fenugreek	6	1	7		
20	Sage	3	12	6		
21	Dill	6	18	5		
22	Nutmeg	1	15	6		
23	Onion	7	20	2		
24	Sesame	5	20	7		

Percent Correct: