ORIGINAL PAPER

Odor span task in dogs (*Canis familiaris*)

Sarah Krichbaum[1] · Bart Rogers[2] · Emma Cox[1] · L. Paul Waggoner[2] · Jeffrey S. Katz[1]

Abstract

Working memory is essential for organisms to solve problems related to their survival and to adapt to changes in their environment. Researchers sought to create a non-human model of working memory that could be used to better understand its predictive value and underlying brain function. Several of these studies were conducted using the *odor span task* (OST) with rodents, and here, we present the first OST with domestic dogs ($n = 6$). The OST is an incrementing non-match-to-sample task in which dogs were presented with both a session novel (S +) and a previously encountered (S −) odor on each trial. A response to the session novel odor was always reinforced. Upon meeting training criterion on sessions with 24 trials or odors to remember, the dogs were tested on the OST with up to 72 odors to remember in the session. All dogs learned the OST and displayed accurate performance ($\geq 79\%$) for the largest set size of 72 odors. In an analysis focused on the effect of intervening odors (i.e., the number of trials since the S − was last encountered), dogs demonstrated above-chance performance for up to eight intervening odors. The implications of these findings are discussed in the context of dog working memory for odors.

Keywords Working memory · Odor span task · Non-match-to-sample · Dog

Introduction

Human working memory is a high-level cognitive process that temporarily stores information for short-term use (Cowan 2017). While specific features of the construct are widely debated, hundreds of studies have shown the importance of working memory in aspects of daily life. Working memory allows organisms to behave appropriately in changing environments, commit goal-directed behaviors and plan for the future (Baddeley 2017). In addition, measures of working memory can reliably predict reasoning skills, learning ability, and other aspects of intelligence (Engle and Kane 2004; lle Lèpine et al. 2005). These measures often address 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 store a limited amount of information (Conway et al. 2005; Scharfen, Jansen, and Holling 2018).

In many years since the identification of the "magic number seven", a number of working memory tasks supported the claim that this process has a very small capacity (Daneman and Carpenter 1980; Kirchner 1958; Miller 1956; Turner and Engle 1989; although see Ma, Husain, and Bays 2014 for an alternative view). 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 and Simon 1973; Scharfen et al. 2018).

Working memory in non-humans is broadly defined as short-term memory for stimuli within a given testing session (Dudchenko 2004; Honig 1978; Olton and Samuelson 1976). Although working memory is typically studied in species such as rats, pigeons, and primates, dogs recently arrived at the forefront of working memory research (Lind et al. 2015). The study of working memory in dogs began with the discovery that aging dogs display similar morphological and cognitive deficits as humans with neurodegenerative disorders (Milgram et al. 1994; Tapp et al. 2003). These findings created an outpour of research that examined how dogs' working memory abilities can inform human and veterinary medicine. The traditional tasks used to study working memory in dogs now span many domains (Bensky et al. 2013).

✉ Sarah Krichbaum
szk0138@auburn.edu

1   Department of Psychology, Auburn University, 226 Thach Hall, Auburn 36849 AL, USA

2   Canine Performance Sciences, Auburn University College of Veterinary Medicine, Auburn, AL, USA

Research on dog working memory is dominated by studies that evaluate the temporal duration of working memory (Bensky et al. 2013; Maclean and Hare 2018). In these studies, dogs 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-position task* (vDNMP) (Adams et al. 2000a; Chan et al. 2002; Head, Rofina, and Zicker 2008; Milgram et al. 1994; Tapp et al. 2003; Salvin et al. 2011; Zanghi et al. 2015). In this task, dogs were presented with a tray that held an object (e.g., a cube) on either the left or right side. Dogs 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). Dogs 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. An incorrect response indicated that the dog was unable to remember the stimulus position after such a delay. While dogs performed well on the vDNMP on delays up to 110 s (Adams et al. 2000a), it is possible that they solved the task by orienting their head or body in the direction of the correct location during the delay which is more indicative of attentional abilities than working memory.

Though informative, limitations notwithstanding, these studies lacked important features of the tasks used to measure working memory in humans. 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. Although duration is an important feature of working memory, a method that defines WMC in dogs is essential to translate findings to working memory in humans. Second, there is evidence that dogs perform better on discrimination tasks that use odor stimuli instead of visual stimuli; therefore, it seemed more appropriate to use dogs dominant sensory modality (Hall, Smith, and Wynne 2013).

An *odor span task* (OST) had the potential to fulfill some of these requirements. The OST is used to examine rats' working memory for odors (Dudchenko, Wood, and Eichenbaum 2000; April et al. 2013). In the study by April et al. (2013), on trial 1, the rat was placed in the center of an arena that contained 18 holes. One hole contained a cup with an odorized lid (S+) and the other 17 holes contained empty cups. A response to the odorized lid (i.e., removing the lid) was reinforced. On trial 2 and on, two holes contained

cups with odorized lids (a session novel odor (S+) and an odor that the rat encountered on a previous trial (S−)). A response to the session novel odor was reinforced. During Experiment 1, the rats completed up to 36 trials per session, meaning that they were required to hold 36 odors in working memory. The results from Experiment 1 showed that the rats responded significantly above chance (50%) for up to 36 odors to remember. In Experiment 2, the researchers sought to find rats' WMC for odors. The rats completed three session types (36 trials, 48 trials, and 72 trials per session) over 2 weeks (1 or 2 sessions per type). The results indicated that rats achieved high accuracy (around 85%) that was significantly above chance for up to 72 odors to remember. When examining accuracy within a session, there was a significant decrease across consecutive trial blocks. However, the rats maintained above-chance accuracy even at the end of a session (around 85% correct; April et al. 2013).

The value of using the OST to measure working memory in dogs was two-fold. First, a direct across-species comparison of the OST could be determined between dogs and rats. Second, working memory could be evaluated in terms of number of odors to remember which may be comparable to working memory measures in humans. Here, we report the first test of working memory in dogs using the OST. Dogs underwent OST training with session lengths of 24 trials (cf. April et al. 2013). We hypothesized that dogs would successfully learn the OST and display high accuracy for up to 24 odors to remember. Next, the session lengths were increased to 36, 48 then 72 trials to further evaluate dogs' WMC. We hypothesized that dogs would display high accuracy for up to 72 odors to remember, similar to what was found in rats.

Methods

Subjects

Six purpose-bred detection dogs (*Canis familiaris)* from the Auburn University Canine Performance Sciences (CPS) program were used in this study. Five of the subjects were Labrador retrievers and one was a Lab mix. The subjects varied in age (M = 3.66 years, range = 2.04–5.06 years) and sex (3 = Females, 3 = Males). Ethical approval for the study was obtained from the Auburn University Institutional Animal Care and Use Committee (Protocol # 2018–3334). 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 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

All training and testing sessions occurred within a 4.5×4 m dog adapted arena in a 18×18 m building at the Canine Performance Science center. Eight, $19 \times 19 \times 19$ cm cinder blocks, with the open end up, were placed on $28 \times 28 \times 18$ cm wooden boxes and arranged in a square formation. The cinderblocks were placed 69 cm apart and spaced with $122 \times 2 \times 30$ cm plywood boards. An additional $109 \times 1 \times 58$ cm plywood board, placed between cinderblocks one and eight, ensured that the dogs' sampled systematically during Phase 1 (see Fig. 1). The holes in each cinderblock were 13×13 cm which allowed the stimulus cans to fit securely in the holes. All experimental sessions were live scored by an experimenter who noted if a response was correct after the blind handler indicated that a response occurred. The handler and experimenters viewed the trials on a monitor located in area C in Fig. 1 and were out of the dogs' view while the dog was in the arena. A Vixia HR70 camera was used to record all sessions.

Stimuli

The stimuli were odorized cotton rounds, 5 cm in diameter, that rested in tins, 6 cm in diameter. The tins were perforated with nine 2 mm holes for odor release. The cotton rounds were odorized by storing them in airtight containers with 28 g of the odor for a minimum of 1 week. In OST pre-training and training, stimuli were selected from a pool of 36 odors. In Exposure or Odor Set Expansion, 36 new odors were added, for a total of 72 odors. All 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 9 cm in diameter, so that they fit securely in the cinderblocks and could be easily handled with metal tongs.

Procedure

Phase 1: Acclimation to the arena The dogs were trained during the first few sessions to respond to the odors. Two experimenters and one dog handler were always present and the sessions consisted of 24 trials. On each trial, a session-novel odor was placed in a random position in the arena by Experimenter 1 (E1). To prevent the dogs from responding based on which locations contained stimulus cans, two empty cans were randomly placed in the arena on each trial. After placing all cans, E1 moved to area A in Fig. 1 out of the dogs' view. Odors were randomly picked without replacement from the pool of 36 odors. The handler released the dog into the arena (from area C in Fig. 1) with the previously trained command "search". The handler remained in area C during all trials to remain out of the dogs' view.

Dogs searched in a counterclockwise fashion from position 1 to 8 (in area B). The order of sampling was controlled by the use of a $109 \times 1 \times 58$ cm plywood board that blocked the dogs' ability to sample in the clockwise direction for the first few sessions. When the dogs were consistently sampling in that direction the board was pushed in so the dogs could freely travel around the arena. The dogs in this study were previously trained to sit in response to a target odor using operant conditioning (see Porritt et al. 2015). Therefore, a sit in front of the position that contained the session novel odor, for 1.5 s, was reinforced. The handler was blind to the

Fig. 1 Dogs began each trial in area **A**. The dogs entered the arena (**B**) through the opening between area A and B, then systematically sampled locations 1–8 in a counterclockwise fashion. The sessions were recorded and live streamed on a monitor (dark bar in **C**)
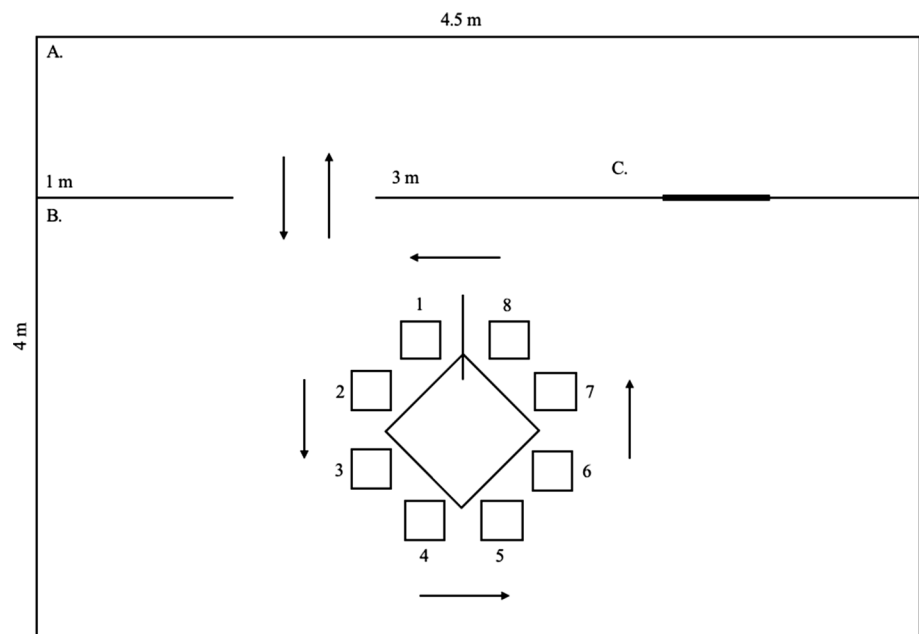
Table 1 Stimuli for OST training (36 odors, top panel) and testing (36 odors, bottom panel)

OST pre-training and 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

Exposure and Odor Set Expansion Test

Amaretto	Apricot	Banana	Black Walnut	Blackberry	Blueberry
Brandy	Butter	Butter Rum	Butterscotch	Caramel	Champagne
Cherry	Chocolate	Clove	Coffee	Cotton Candy	Eggnog
English Toffee	Grape	Honey	Lemon	Lime	Maple
Marjarom	Marshmallow	Peach	Peanut Butter	Pecan	Pineapple
Pistachio	Pumpkin	Raspberry	Rootbeer	Strawberry	Tangerine

position of the session novel odor and used a live monitor to indicate when the dog responded to a position. A second experimenter (E2) remained in area C and either marked the dog's response with a click if the dog responded to the session novel odor or withheld the click if the dog responded to any other position, as indicated by the handler with a hand raise. The click cued the dog to exit area B to area C and receive the reinforcer (e.g., ball). Once the dog was successfully sitting in front of the odors at all eight locations, Phase 2 began.

Phase 2: OST pre-training During this phase, dogs always completed 24 trials in a session (collected on a single day). A session could occur in 3 blocks of 8 trials, 2 blocks of 12 trials, or 1 block of 24 trials with 10 min between blocks. The blocking was determined by the dogs' 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 were 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 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. The dogs were sent into the arena the same as in Phase 1. 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 reinforcer in area C. On every subsequent trial, the dog was presented with a two-choice discrimination between a previously encountered odor from that session (S−), selected from one of the last five trials to not increase memory load, and a session novel odor for that session (S+). The S+ odors were randomly selected from the pool of 36 odors without replacement. Both odors (S+ and S−) were placed in random locations

on the wheel. The dog received the reinforcer for responding to the S+.

Two types of correction procedures were used in this phase. First, a time-out procedure was implemented so that if the dog responded to the previously encountered odor (S−), the handler called the dog to exit the arena to area C with a previously trained command "come", held the dog for 10 s, then released the dog back in to the arena. The trial was repeated until the dog responded to the S+ without a response to the S−. Second, a wait-out (i.e., autocorrection) procedure was implemented such that if the dog responded to the previously encountered odor (S−), the dog remained in the arena until responding to the S+. The type of correction procedure used depended on individual dog need and they could receive both types within a single session. Once a performance criterion of a minimum of ten sessions with at least 75% correct in a session was met, dogs advanced to Phase 3.

Phase 3: OST training All 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, six 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 for that session. 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 that dogs could reject an odor due to smelling themselves on it. 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% correct for two consecutive sessions (with no time-out 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% correct on a session of 24 trials (deemed maintenance sessions).

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 Odor Set Expansion Test in Phase 5 to determine if exposure to the entire odor set had an effect on performance. It was possible that prior exposure to the odors could impact performance in the Odor Set Expansion Test. For example, rapidly expanding the number of odors in Phase 5 could negatively impact performance due to the increase in novel odors. If so, dogs in Phase 4 would perform better than dogs not exposed to the odors in Phase 5. In contrast, novel odors in Phase 5 could be more sailent and in in turn facilitate performance. Hence, to determine if novelty was an important variable, the other four dogs did not undergo the Exposure phase before moving to Phase 5. In Phase 4, each session was 24 trials and consisted of nine OST training sessions. Its purpose was to expose the dog to all 72 odors that would be presented during Phase 5. Therefore, the 72 stimuli were balanced for number of presentations as the S + 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% correct on a session of 24 trials (deemed maintenance sessions). Upon meeting this criterion, the dogs advanced to Phase 5.

Phase 5: Odor Set Expansion Test All the dogs completed the Odor Set Expansion Test. The sessions occurred the same as the OST training 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, for example, the dogs were required to remember up to 72 odors in the 72-trial session. The odor stimuli were selected and set in the arena in the same way as Phase 3 with the addition of the 36 new odors (lower half of Table 1). Each dog completed two sessions of each testing condition. During the first week of testing, the sessions were conducted in ascending order (i.e., 36, 48, then 72 odors), similar to the design used with rats (April et al. 2013). During the second week, that order was repeated.

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 a sit in front of the S + without a sit in front of the previously encountered odor (S −) or an empty can. For comparison to rats, percent correct was analyzed across blocks of 12 trials (April et al. 2013). The main variable of interest was percent correct for each odor set size. However, visits (i.e., the number of trials that the dogs' nose came within 5 cm of the S − odor) were also recorded. Additional post hoc analyses focused on the number of intervening odors (i.e., the number of trials since an S − was last encountered as either an S + or an S −) and the number of spaces between the S + and S − (i.e., the number of empty stimulus positions between the S + and S −) on percent correct.

Data sheets were created and live scored by E2 for all OST pre-training, OST training, Exposure, and the Odor Set Expansion Test sessions. 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 for each dog.

Results

Training The number of total sessions in each phase varied due to the number of sessions required for 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 the dogs acquired the task and met the OST training performance criterion of two consecutive sessions of 84% 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% correct, as determined by one-sample t tests $t(5) = 1.91$, $p = 0.12$ and $t(5) = 1.41$, $p = 0.22$, respectively.

Odor Set Expansion Test Figure 2 shows the average and individual dog performance across the three odor set

Table 2 Number of sessions that each dog completed for different phases of training and testing in OST

Subject	OST pre-training	OST training	Maintenance	Exposure	OSET	Total
Beaufort	13	32	16	9	6	76
Jolly	16	16	19	9	6	66
Kylee	16	20	2	–	6	44
Nessie	20	10	8	–	6	44
Quail	26	15	6	–	6	53
Vera	12	20	1	–	6	39

OST odor span task, *OSET* Odor Set Expansion Test

Fig. 2 Average and individual dog performance shown as overall percent correct for each odor set size (36 odors: black bars, 48 odors: dark gray bars, 72 odors: light gray bars). Error bars represent SEM
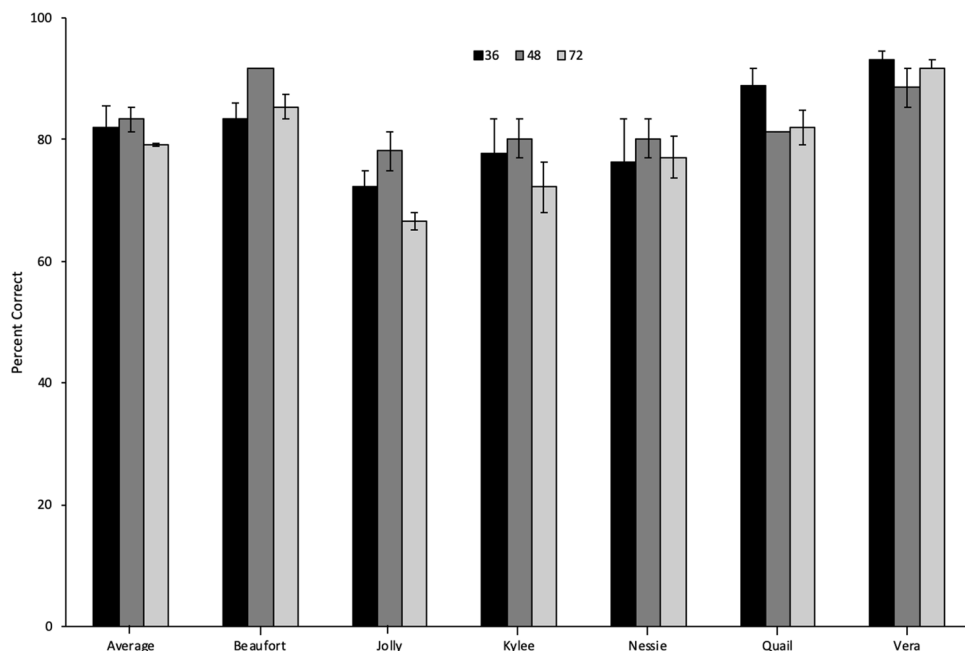

sizes. Overall, the dogs displayed comparable accuracy for the 36, 48, and 72 odor set sizes, as confirmed by a two-way repeated measures ANOVA of set size (36, 48, 72) and replication (first, second) which revealed no main effect of replication $F(1, 5) = 0.07$, $p = 0.80$, $\eta_p^2 = 0.02$, or set size $F(2, 10) = 1.97$, $p = 0.19$, $\eta_p^2 = 0.28$, and no significant interaction $F(2, 10) = 2.06$, $p = 0.18$, $\eta_p^2 = 0.29$. Given that performance on each set size was not influenced by replication, the data within each set size were averaged for the remaining analyses. The dogs exhibited above-chance accuracy (50%) on set sizes of 36, 48, and 72 odors, as determined by one-sample t tests, $t(5) = 25.08$, $p < 0.001$, $t(5) = 37.35$, $p < 0.001$, $t(5) = 21.23$, $p < 0.001$, respectively. There was

no difference in Odor Set Expansion Test accuracy between the dogs (Beaufort and Jolly: $M = 79.58$, SEM $= 2.71$) that experienced the entire odor set prior to testing in Phase 4 and the dogs (Kylee, Nessie, Quail, and Vera: $M = 82.44$, SEM $= 0.95$) that did not experience the entire set prior to testing, as determined by an independent-sample t test, $t(4) = 0.93$, $p = 0.69$. Therefore, data from all the dogs were included in the analyses as a single group.

Within session performance for each odor set size in 12-trial blocks can be seen in Fig. 3. The dogs displayed a similar decrease in accuracy across trial blocks for set sizes of 36, 48 and 72. A series of ANOVAs supported these findings. A two-way repeated measures ANOVA for trial block

Fig. 3 Average percent correct across trials (12-trial blocks) for each odor set size (36 odors: black circles; 48 odors: dark gray circles; 72 odors: light gray circles). Error bars represent SEM and the dashed line represents chance performance
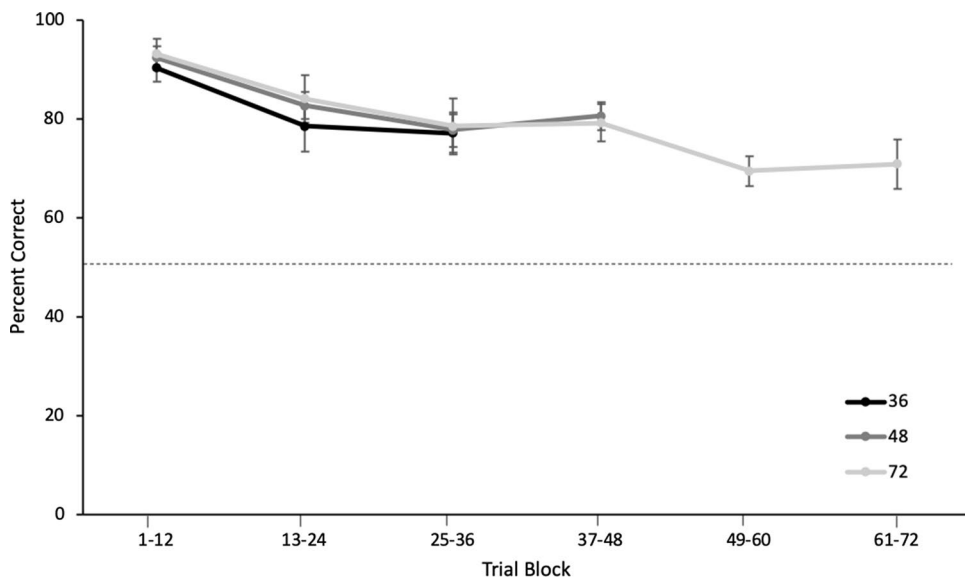
(1–3) and set size (36, 48, 72) revealed a main effect of trial block $F(2, 10) = 30.40$, $p < 0.001$, $\eta_p^2 = 0.89$, no main effect of set size $F(2, 10) = 0.81$, $p = 0.47$, $\eta_p^2 = 0.14$, and no interaction $F(4, 20) = 0.202$, $p = 0.93$, $\eta_p^2 = 0.04$. Similarly, a two-way repeated measures ANOVA for trial block (1–4) and set size (48, 72) revealed a main effect for trial block $F(3, 15) = 19.50$, $p < 0.001$, no main effect for set size $F(1, 5) = 0.02$, $p = 0.90$ $\eta_p^2 = 0.004$, and no interaction $F(3, 15) = 0.17$, $p = 0.91$, $\eta_p^2 = 0.03$. A one-way ANOVA for trial block (1–6), for the odor set size of 72, also revealed a main effect of trial block $F(5, 25) = 13.17$, $p < 0.001$, $\eta_p^2 = 0.80$. Additionally, the decrease in accuracy across trial blocks was similar for all set sizes as supported by a one-way repeated measures ANOVA that compared the slopes for each dog at each set size, $F(2, 10) < 1$, $p = 0.881$, $\eta_p^2 = 0.025$. Importantly, the dogs maintained above-chance accuracy for all trial blocks for each set size, as confirmed by one-sample t tests, $ps < 0.005$.

Odor controls and spacing Odor control ($M = 84.26$, SEM $= 5.40$) and regular ($M = 80.67$, SEM $= 2.81$) trials were compared across testing and were not different, as supported by a paired-samples t test, $t(5) = -1.02$, $p = 0.35$, ruling out scent marking. The dogs scored significantly above chance on regular and odor control trials, as confirmed by one-sample t tests, $t(5) = 10.92$, $p < 0.001$ and $t(5) = 6.29$, $p = 0.001$, respectively. Additionally, the effect of number of spaces between the S+ and S− did not influence accuracy, as disclosed by a one-way repeated measures ANOVA on accuracy $F(3, 15) = 0.29$, $p = 0.84$, $\eta_p^2 = 0.05$.

Intervening odors Figure 4 depicts performance as a function of intervening odor block during the six testing sessions, averaged across dogs, in blocks of two intervening odors. 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 visited the S− odor on over 50% of the testing trials ($M = 60.80$, SEM $= 3.36$) as supported by a one-sample t test, $t(5) = 3.22$, $p = 0.02$. It is important to note that the number of trials analyzed in each block was not counterbalanced. Specifically, there were 269, 133, 103, 77, 87, and 67 trials, respectively, across the six intervening odor blocks. 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 intervening odor block $F(5, 25) = 3.34$, $p = 0.02$, $\eta_p^2 = 0.40$. Although there was a decrease in accuracy across number of intervening odors, the dogs displayed significantly above-chance performance for intervening odors of 1–2, 3–4, 5–6 and 7–8 as confirmed by one-sample t tests, $t(5) = 10.02$, $p < 0.001$, $t(5) = 6.83$, $p = 0.001$, $t(4) = 9.01$, $p = 0.001$, $t(5) = 3.18$, $p = 0.038$, respectively. One dog was an outlier (two standard deviations below the mean) and was removed from the analysis comparing intervening odors of 5–6 to chance.

Discussion

All dogs learned the OST and performed at high levels (> 84%) on the last two OST training sessions before proceeding to testing with increased odor set sizes. These results supported the hypothesis that dogs could learn the two-choice non-match-to-sample-task. This high level of performance continued throughout testing and was not affected by the size of the stimulus set (36, 48, 72), exposure to the set of odors prior to testing, and was replicated. 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 eight intervening odors during testing. Other variables, including spacing of

Fig. 4 Average percent correct across number of intervening odors (two intervening odor blocks) for all test sessions (*indicates significant one-sample t test, $p < 0.05$). Error bars represent SEM and the dashed line represents chance performance
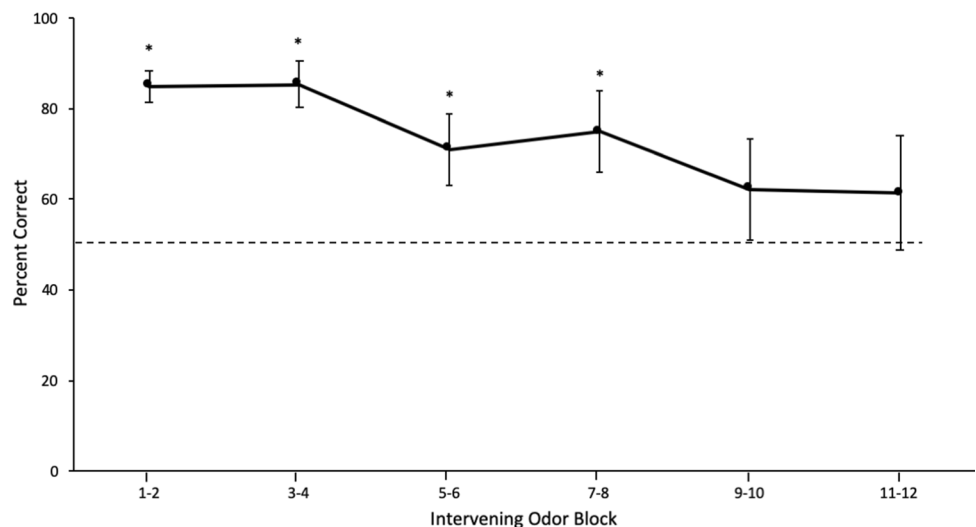
the S + and S − and odor controls (ruling out scent marking), did not affect accuracy during testing. These results are discussed in regard to different views of working memory.

As expected, dogs performed similar to rats when within session accuracy was examined for each set size (cf. April et al. 2013). Dogs showed a similar 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 odor set size. Similar to what was found in rats, the slopes of all three functions were shallow and the dogs remained well above (73.83%) chance on the last trial block for the set size of 72 odors. It is possible that overall accuracy was comparatively high in dogs, like rats, because they have a WMC for up to 72 odors. While rats and dogs were similar in their overall performance, there was a difference in their sampling behavior. The dogs sampled the S − significantly more than chance. In other words, if they encountered the S + first, they typically left it and proceeded to the S − before responding. These results contrast what was reported in rats, as rats rarely proceeded to the S − before responding (Galizio et al. 2016). Therefore, it is reasonable to conclude that dogs, unlike rats, preferred to sample both odors before responding and suggests a possible difference in how they solve the task. Additional species should be tested on versions of the span task that implement different stimulus modalities (e.g., a visual span task in pigeons) to determine if these results are specific to odor memory.

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, as presented based on number of odors in a session, does not measure WMC (April et al. 2013). First, the Odor Expansion Test sessions that consisted of 72 trials took an average of 45 min to complete. Therefore, a discussion of the high accuracy at the end of the session is in direct opposition of the working memory hypothesis which suggests memory of short duration. Second, the dogs could have responded based on the process known as familiarity. Familiarity works parallel to working memory but rather encodes if a stimulus has been encountered previously (e.g., Basile and Hampton 2013; Pañoz-Brown et al. 2016; Yonelinas 2002). For example, the dogs could have avoided the S − due to a sense of familiarity (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 and Hampton 2013). Future work should seek to tease apart these constructs in an effort to better understand the mechanism of working memory in dogs (cf. Basille and Hampton 2013; Brady and Hampton 2018).

A new way to determine WMC in the OST could be to assess the number of intervening odors or 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 trials ago or second to last (Redick and Lindsey 2013). Although n-back tasks are a traditional way to study WMC, WMC and working memory duration are typically confounded. The task assesses WMC by increasing n-back but it also influences working memory duration by increasing the amount of time since the item was last encountered. In any case, 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.

Dogs displayed significantly above-chance accuracies for up to 7–8 intervening odors. This finding suggests that the OST may be used as a measure of WMC for odors in dogs. If the dogs were strictly using familiarity, there would not be an effect of intervening odors, because familiarity is thought to be infinite. Alternatively, it could be that the dogs have a familiarity criterion that can change with experience (Wright 2012). With a lax criterion, the function in Fig. 4 would be flat, but as the criterion becomes stricter the slope of the function would be steeper. As discussed, this analysis cannot dissociate WMC from working memory duration, therefore, future research varying the interval between trials and odor set size will be important to further unravel the processes the underlie memory in the OST (Bratch et al. 2016). In addition, it is important to note that the number of observations in each block of intervening odors was not counterbalanced in the present study, as the S − on each trial was randomly selected from the odors previously encountered in the session. Specifically, the number of observations decreased with increasing number of intervening odors. Future work should implement a counterbalancing technique to create an equal number of observations across the number of intervening odors (cf. Wright Katz and 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 scent marking. During testing, the dogs showed no difference in accuracy on odor control and regular 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 the 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 the odor source regardless of whether they could smell the odors upon entering the arena. It is possible that the dogs' previous training enabled their ability to accurately locate and discriminate the odors even when in close proximity. Future research should explore whether training (e.g., scent detection or nose work) facilitates performance on the task.

Understanding working memory in detection dogs may have implications for this population. Detection dogs are utilized for a wide range of scent detection tasks, 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 drop-out rate in training programs (Cobb et al. 2015; Maejima et al. 2007). There is a clear need for improving detection dog selection and training to increase the number of dogs successfully placed in service. Because of the complex cognitive processes involved in many detection dog roles (Maclean and Hare 2018), researchers have expressed the need to increase our understanding of cognition in relation to the services they provide (Troisi et al. 2019). 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 (Maclean and Hare 2018).

Exploring the OST in other dog populations including pet dogs is important to determine if the high level of performance by detection dogs in the current study is due to their unique odor training or a mechanism that is common to all dogs. Based on studies that used the OST in experienced and naïve rats, it is reasonable to hypothesize that dogs with little or no odor training would eventually learn the OST (April et al. 2013; Dudchenko et al. 2000). However, there are likely to be population or breed differences in rate of acquisition and overall performance due to varying olfactory ability. For example, there is evidence that dog breeds selected for detection work outperform non-scent and short-nosed pet dog breeds on a natural detection task (Polgár et al. 2016) suggesting that specific breeds were selected for their olfactory abilities. It is important to pinpoint how breed differences in cognition contribute to the selection and ultimate success in different working roles (Bray et al. 2015; MacLean and Hare 2018).

The OST as a model of working memory in dogs may hold value in both veterinary and human medicine. For example, the aging dog brain shows similarities to humans with neurodegenerative disorders (Tapp et al. 2003). In addition, veterinarians have discovered Cognitive Dysfunction Syndrome (CDS) in dogs that increase with age and share features with human aging diseases (Bain et al. 2001; Studzinski et al. 2006). 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 identify dogs that are at risk for or in early stages of CDS. This would increase the number of dogs that are diagnosed and properly treated, while serving as a valuable model for similar human diseases. 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.

The present study represents the first results of dogs' ability to perform the OST demonstrating memory for odors over a short period of time, with several key findings. First, the dogs displayed high accuracy for up to 72 trials or odors to remember which is indicative of their ability to not only learn non-match-to-sample with odors but also to make 72 consecutive 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 eight intervening odors since the $S-$ was encountered which may be an indicator of WMC for odors. Although it is necessary to dissociate the processes of familiarity and working memory, the OST may have important value in the training and selection of detection dogs as well as informing veterinary and human medicine.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards This research complied with the current laws of the United States of America and was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at Auburn University.

References

Adams B, Chan A, Callahan H, Milgram NW (2000) The canine as a model of human cognitive aging: recent developments. Progr Neuro-Psychopharmacol Biol Psychiatry 24(5):675–692

April LB, Bruce K, Galizio M (2013) The magic number 70 (plus or minus 20): variables determining performance in the rodent odor span task. Learn Motiv 44(3):143–158

Baddeley AD (2017) The concept of working memory: A view of its current state and probable future development. Exploring Working Memory Routledge, pp 99–106

Bain MJ, Hart BL, Cliff KD, Ruehl WW (2001) Predicting behavioral changes associated with age-related cognitive impairment in dogs. J Am Vet Med Assoc 218(11):1792–1795

Basile BM, Hampton RR (2013) Dissociation of active working memory and active working memory and passive recognition in rhesus monkeys. Cognition 126(3):391–396

Bensky MK, Gosling SD, Sinn DL (2013) The world from a dog's point of view: a review and synthesis of dog cognition research. Adv Study Behav 45:209–406

Brady RJ, Hampton RR (2018) Nonverbal working memory for novel images in rhesus monkeys. Curr Biol 28(24):3903–3910

Bratch A, Kann S, Cain JA, Wu JE, Rivera-Reyes N, Dalecki S, Arman D, Dunn A, Cooper S, Corbin HE, Doyle AR, Pizzo MJ, Smith AE, Crystal J (2016) Working memory systems in the rat. Curr Biol 26(3):351–355

Bray EE, MacLean EL, Hare BA (2015) Increasing arousal enhances inhibitory control in calm but not excitable dogs. Anim Cogn 18(6):1317–1329

Chan AD, Nippak P, Murphey H, Ikeda-Douglas CJ, Muggenburg B, Head E, Cotman CW, Milgram NW (2002) Visuospatial impairments in aged canines (Canis familiaris): the role of cognitive-behavioral flexibility. Behav Neurosci 116(3):443

Chase WG, Simon HA (1973) Perception in chess. Cogn Psychol 4(1):55–81

Cobb M, Branson N, McGreevy P, Lill A, Bennett P (2015) The advent of canine performance science: offering a sustainable future for working dogs. Behav Proc 110:96–104

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW (2005) Working memory span tasks: A methodological review and user's guide. Psychon Bull Rev 12(5):769–786

Cowan N (2017) The many faces of working memory and short-term storage. Psychon Bull Rev 24(4):1158–1170

Daneman M, Carpenter PA (1980) Individual differences in working memory and reading. J Verbal Learn Verbal Behav 19(4):450–466

Dudchenko PA (2004) An overview of the tasks used to test working memory in rodents. Neurosci Biobehav Rev 28(7):699–709

Dudchenko PA, Wood ER, Eichenbaum H (2000) Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. J Neurosci 20(8):2964–2977

Engle RW, Kane MJ (2004) Executive attention, working memory capacity, and a two-factor theory of cognitive control. Psychol Learn Mot 44:145–200

Galizio M, April B, Deal M, Hawkey A, Panoz-Brown D, Prichard A, Bruce K (2016) Behavioral pharmacology of the odor span task: Effects of flunitrazepam, ketamine, methamphetamine and methylphenidate. J Exp Anal Behav 106(3):173–194

Hall NJ, Smith DW, Wynne CD (2013) Training domestic dogs (Canis lupus familiaris) on a novel discrete trials odor-detection task. Learn Motiv 44(4):218–228

Head E, Rofina J, Zicker S (2008) Oxidative stress, aging, and central nervous system disease in the canine model of human brain aging. Vet Clin 38(1):167–178

Honig WK (1978) Studies of working memory in the pigeon. Cognit Processes Anim Behav 211–248

Kirchner WK (1958) Age differences in short-term retention of rapidly changing information. J Exp Psychol 55(4):352

IleLépine R, Parrouillet P, Camos V (2005) What makes working memory spans so predictive of high-level cognition. Psychon Bull Rev 12(1):165–170

Lind J, Enquist M, Ghirlanda S (2015) Animal memory: A review of delayed matching-to-sample data. Behav Proc 117:52–58

Ma WJ, Husain M, Bays P (2014) Changing concepts in working memory. Nat Neurosci 17:347–356

Maejima M, Inoue-Murayama M, Tonosaki K, Matsuura N, Kato S, Saito Y, Weiss A, Murayama Y, Ito SI (2007) Traits and genotypes may predict the successful training of drug detection dogs. Appl Anim Behav Sci 107(3):287–298

MacLean EL, Hare B (2018) Enhanced selection of assistance and explosive detection dogs using cognitive measures. Front Vet Sci 5:236

Milgram NW, Head E, Weiner E, Thomas E (1994) Cognitive functions and aging in the dog: acquisition of nonspatial visual tasks. Behav Neurosci 108(1):57

Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 101(2):343–352

Olton DS, Samuelson RJ (1976) Remembrance of places passed: spatial memory in rats. J Exp Psychol 2(2):97

PañozBrown D, Corbin H, Dalecki S, Gentry M, Brotheridge S, Sluka C, Wu J, Crystal J (2016) Rats remember items in context using episodic memory. Current Biol 26(20):2821–2826

Polgár Z, Kinnunen M, Újváry D, Miklósi Á, Gácsi M (2016) A test of canine olfactory capacity: comparing various dog breeds and wolves in a natural detection task. PLoS ONE 11(5):e0154087

Porritt F, Shapiro M, Waggoner P, Mitchell E, Thomson T, Nicklin S, Kacelnik A (2015) Performance decline by search dogs in repetitive tasks, and mitigation strategies. Appl Anim Behav Sci 166:112–122

Redick TS, Lindsey DR (2013) Complex span and n-back measures ofworking memory: a meta-analysis. Psychon Bull Rev 20(6):1102–1113

Salvin HE, McGreevy PD, Sachdev PS, Valenzuela MJ (2011) Growing old gracefully—Behavioral changes associated with "successful aging" in the dog, Canis familiaris. J Vet Behav 6(6):313–320

Scharfen J, Jansen K, Holling H (2018) Retest effects in working memory capacity tests: A meta-analysis. Psychon Bull Rev 25:2175–2199

Studzinski CM, Christie LA, Araujo JA, Burnham WM, Head E, Cotman CW, Milgram NW (2006) Visuospatial function in the beagle dog: an early marker of cognitive decline in a model of human aging and dementia. Neurobiol Learn Memory 86(2):197–204

Tapp PD, Siwak CT, Estrada J, Head E, Muggenburg BA, Cotman CW, Milgram NW (2003) Size and reversal learning in the beagle dog as a measure of executive function and inhibitory control in aging. Learn Mem 10(1):64–73

Troisi CA, Mills DS, Wilkinson A, Zulch HE (2019) Behavioral and cognitive factors that affect the success of scent detection dogs. Comp Cognit Behav Rev 14:51–67

Turner ML, Engle RW (1989) Is working memory capacity task dependent? J Mem Lang 28(2):127–154

Yonelinas AP (2002) The nature of recollection and familiarity: A review of years of 30 years of research. J Mem Lang 46(3):441–517

Wright AA (2012) Memory processing. In: Zentall TR, Wasserman EA (eds) The Oxford Handbook of Comparative Cognition. Oxford University Press, New York, pp 239–260

Wright AA, Katz JS, Ma WJ (2012) How to be proactive about interference: Lessons from animal memory. Psychol Sci 23(5):453–458

Zanghi BM, Araujo J, Milgram NW (2015) Cognitive domains in the dog: independence of working memory from object learning, selective attention, and motor learning. Anim Cognit 18(3):789–800