ORIGINAL PAPER

Adaptive spatial working memory assessments for aging pet dogs

Joshua Van Bourg[1] · Rachel Gilchrist[1] · Clive D. L. Wynne[1]

Abstract

Assessments for spatial working memory (SWM) in pet dogs that can detect age-related cognitive deficits in a single session may aid in diagnosing canine dementia and may facilitate translational research on Alzheimer's disease in humans. Adaptive testing procedures are widely used in single-session assessments for humans with diverse cognitive abilities. In this study, we designed and deployed two up-down staircase assessments for SWM in which 26 pet dogs were required to recall the location of a treat hidden behind one of two identical boxes following delays of variable length. In the first experiment, performance tended to decline with age but few dogs completed the test ($n = 10$). However, all of the dogs that participated in the second experiment ($n = 24$) completed the assessment and provided reliable evidence of learning and retaining the task. Delay length and age significantly predicted performance supporting the validity of this assessment. The relationships between age and performance were described by inverted U-shaped functions as both old and young dogs displayed deficits in weighted cumulative-scores and trial-by-trial performance. Thus, SWM in pet dogs may develop until midlife and decline thereafter. Exploratory analyses of non-mnemonic fixation strategies, sustained engagement, inhibitory control, and potential improvements for future SWM assessments which adopt this paradigm are also discussed.

Keywords Dog · Cognitive decline · Aging · Development · Spatial working memory · Staircase methods

Introduction

As in humans, the cognitive abilities of dogs may decline with age. For some, these impairments are minor and only present late in life (Chapagain et al. 2018; Head 2001). For others, cognitive decline begins early in life and progresses more rapidly, ultimately leading to severe impairments (Adams et al. 2000a; Head 2001). As with Alzheimer's disease (AD) in humans (Kensinger et al. 2003; Simone and Baylis 1997), age-related cognitive deficits (ARCD) in dogs are characterized by the deterioration of recent memory and other executive functions including selective attention, behavioral inhibition, and concept learning (Adams et al. 2000a; Head 2013). Although the pathogenesis of ARCD is unclear, both ARCD and AD are correlated with the

development of neuritic beta-amyloid plaques (Head 1998; Vite and Head 2014). In clinical settings, the development of behavioral impairments in elderly pet dogs is often referred to as cognitive dysfunction syndrome (CDS) (Landsberg et al. 2012; Szabó et al. 2016). As with AD, CDS is characterized by increased anxiety, nocturnal activity, agitation, and disorientation (Madari et al. 2015; Landsberg et al. 2012; Fast et al. 2013).

This cognitive, behavioral, and physiological overlap suggests that dogs may provide a strong animal model for translational AD research as well as a naturally occurring alternative to transgenic rodent and invertebrate models (Araujo et al. 2017). In addition, dogs are highly tractable, widely accessible, and share a number of anatomical similarities with humans but have shorter lifespans (Gilmore and Greer 2015; Head 2013; Kaeberlein et al. 2016). Furthermore, research on ARCD in pet dogs living with elderly owners may provide insight into the environmental correlates of AD (Kaeberlein et al. 2016).

ARCD in dogs are domain specific and are more readily detected with complex tasks that require coordination of multiple cognitive functions (i.e., executive functions). For example, older beagles are slower to acquire, and commit

✉ Joshua Van Bourg
 joshvanbourg@asu.edu

[1] Department of Psychology, Arizona State University, Tempe, AZ, USA

more errors in, tests for egocentric spatial learning, oddity and size discrimination, and reversal learning but do not display deficits in simple visual discrimination or procedural learning tasks (see Head 2013 for review). Different studies have produced different estimations for the onset age of dementia in dogs but most suggest that neurological, cognitive and behavioral changes begin in midlife (for review, see Chapagain et al. 2018; Szabó et al. 2016). Importantly, sensorimotor abilities and some executive functions including attention and working memory may continue to develop after adolescence in dogs (Wallis et al. 2014; Watowich et al. 2020). Tests for spatial working memory (SWM), "the process of maintaining a limited amount of spatial information in an active representation for a short period of time so that it is available for use" (Adams et al. 2000b, p. 48), are perhaps the most useful assessments for ARCD in dogs (Head et al. 2013). Moreover, SWM deficits are correlated with cortical atrophy (Rofina et al. 2006), reduced neurogenesis (Siwak-Tapp et al. 2007), and old age in colony beagles (Adams et al. 2000b; Chan et al. 2002; Tapp et al. 2003b; Studzinski et al. 2006).

Research on ARCD in pet dogs is currently limited by a lack of reliable cognitive assessments which can be administered in a single session using cheap and accessible materials and without requiring extensive behavioral shaping (Piotti et al. 2017). SWM assessments for colony beagles implement stringent training criteria and yield reliable performance estimates but require multiple shaping, training, and testing sessions. Indeed, established tests for SWM require weeks or months of daily sessions (e.g., Adams et al. 2000b; Milgram et al. 1999). In contrast, existing SWM assessments for pet dogs (González-Martínez et al. 2013; Piotti et al. 2017) can be completed in a single session but do not implement training criteria or other important controls. Ultimately, valid and reliable single-session assessments for SWM are needed to better understand ARCD in pet dogs.

Such assessments may help to clarify the relationships between age-related cognitive deficits, behavioral changes, and neurological changes in pet dogs. Used in combination with physiological tests and questionnaires, cognitive assessments may also aid in the diagnosis of CDS (Wallis et al. 2016), thereby facilitating veterinary care for millions of pet dogs (for prevalence estimates, see Azkona et al. 2009; Salvin 2010). Moreover, physicians use a number of tools to diagnose AD including behavioral questionnaires, psychometric assessments, physiological assays, and neurological imaging. In contrast, veterinarians primarily diagnose CDS using only owner reports and questionnaires, the validity and consistency of which remain controversial (Szabó et al. 2016, 2018).

Adaptive testing procedures (ATPs) accommodate a wide range of abilities within a single assessment using a response-dependent testing strategy to tailor each test to the individual test-taker (Wainer and Lewis 1990). Relative to traditional tests which present tasks spanning a broad range of difficulty to all test-takers, ATPs improve efficiency by omitting tasks that are far from an individual's performance threshold—tasks that are too challenging or too simple (Marinissen et al. 2010; Watson and Fitzhugh 1990). ATPs are primarily used to estimate maximal ability, peak performance, or cumulative performance. For example, the Graduate Management Admission Test®, a computerized adaptive test that determines the difficulty of future questions based on the outcomes and difficulty of preceding questions, produces a weighted cumulative score which awards more points for correct answers to difficult questions.

Up–down staircase methods, in which an individual may advance or regress through a sequential experimental design, have been widely used to measure performance thresholds in humans and non-human animals (Cornsweet 1962; Ehrenstein and Ehrenstein 1999). For example, staircase procedures are commonly used in hearing tests to rapidly determine the highest frequency that is audible to the test-taker. In each trial, the proctor plays a recording of a short tone and the test-taker indicates whether she hears this tone. When this occurs, the proctor plays a tone of a higher frequency in the following trial. Otherwise, the proctor plays a tone of lower frequency in the following trial.

In staircase assessments, test-takers oscillate between completing tasks which fall within the range of their abilities and failing tasks which exceed their abilities. Thus, staircase methods can reliably estimate performance thresholds without stringent progression criteria (i.e., criteria for "stepping-up" in the staircase). Given that task difficulty is additive, staircase procedures can also rapidly estimate ability by measuring performance within ranges of task difficulty.

The purpose of this study was to design an adaptive test for SWM which can reliably and rapidly detect ARCD in pet dogs. In two experiments, we designed and deployed assessments in which 26 pet dogs were required to recall the location of a treat hidden behind one of two boxes following a delay (retention period) of variable length. The first assessment used a transformed staircase procedure (Kaernbach 1991) which required the dog to make consecutive correct choices to progress into trials with longer delays, and implemented zero-delay correction trials following incorrect choices. To address an apparent ceiling effect and to reduce the rate of fail-out, the second assessment used a transformed and weighted staircase procedure (Kaernbach 1991) in which the magnitude of change in the delay was also performance-dependent (i.e., step sizes varied).

To test the validity of the modified assessment in the second experiment, we examined whether each dog learned and followed the rules of the task, age predicted performance, and performance was delay-dependent. We then conducted supplementary analyses of the dog's head
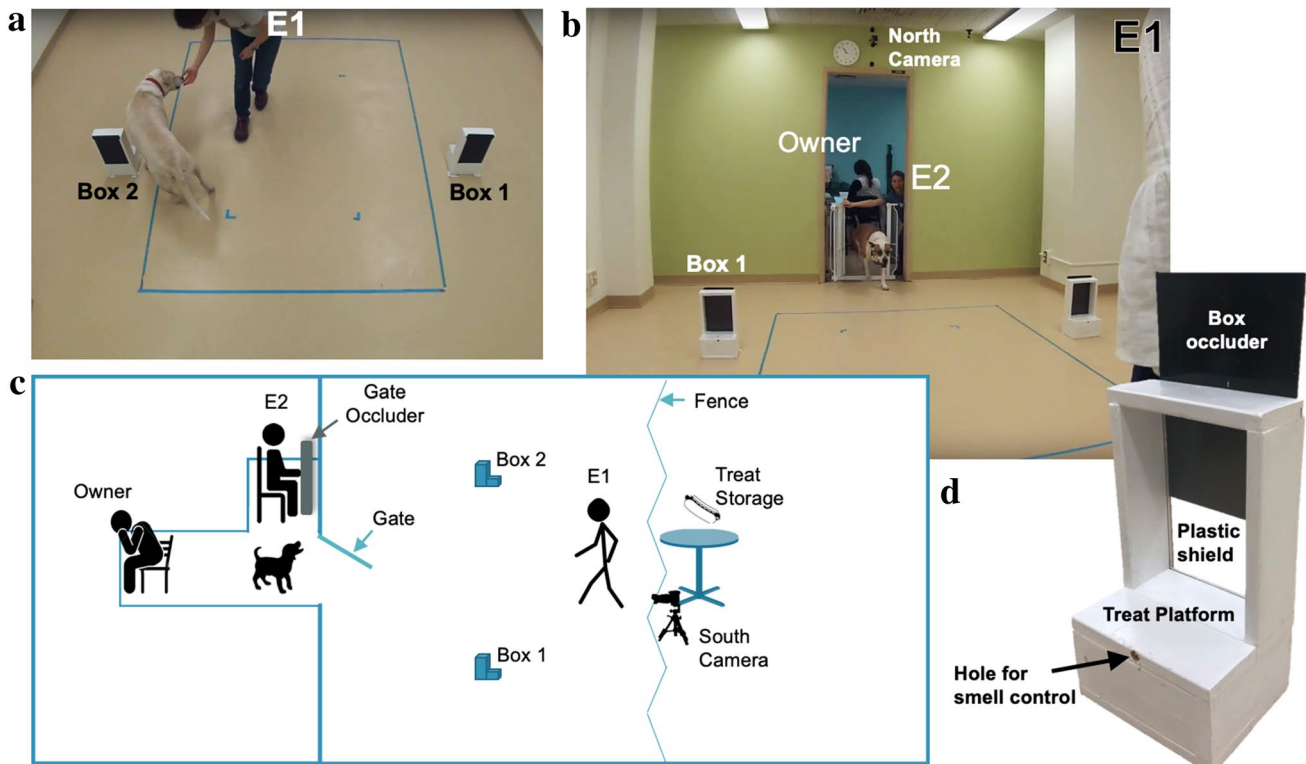
Fig. 1 Schematic of the experimental layout. **a** View from north camera. **b** View from south camera. **c** Layout of the rooms in which the experiments were conducted. **d** Picture of a treat box

and body orientations to explore the processes through which dogs stored and maintained information in SWM and to assess whether the dog's ability to remain engaged with the task could account for age effects observed in the primary analyses. Finally, we conducted two exploratory analyses to inform the design of future assessments which could adopt this paradigm. Specifically, we examined whether this experimental design may be used to detect age-related deficits in inhibitory control, and whether a shorter, fixed-delay procedure may be used to detect age-related SWM deficits.

Method

Subjects

Twenty-six pet dogs (13 of each sex) of various breeds and ages (17–181 months, mean = 89.82) were tested (see supplementary material S1). Given that the age estimation of an adult dog may not be reliable, only dogs seen by a veterinarian before reaching 1 year of age were included in this study. All owners volunteered their dogs to participate.

Materials and layout

One-meter tall collapsible, plastic fencing was used to create a 5 by 5 m testing area and a 2 by 1 m holding area in an adjacent room (Fig. 1). A 1 by 1 m doorframe-mounted dog gate was used to restrict the dog to the holding area without blocking visibility into the testing area. To prevent visibility, a $1.5 \times 1.0 \times 0.04$ m foam board was placed against the dog gate. Throughout each session, a demonstrator (E1) remained in the testing area, an assistant (E2) and the owner remained in the holding area, and the dog moved between the two areas. To minimize distraction while preventing separation anxiety, the owner sat in a chair opposite the testing room facing away from the experimental area, reading silently and ignoring the dog. To control access and visibility into the testing room while minimizing interactions with the dog, E2 sat in a chair directly adjacent to the doorway facing the testing room.

Treats were hidden in one of two identical boxes placed on the ground in the testing area (Fig. 1). These boxes were spaced 2 m apart and 2.5 m from the starting location of both the dog and E1. Each box consisted of a $0.2 \times 0.2 \times 0.1$ m base, upon which the experimenter placed food treats, and a $0.1 \times 0.2 \times 0.2$ m shield, which could be configured to

Table 1 Example of training progression in Experiment I

Trial	1	2	3	4	5	6	7	8	9	10	11	12
Level	1	1	2	2	2	2	3	3	3	3	3	3
Choice	**L**	**R**	**L**	*L*	**L**	**R**	*L*	*L*	**L**	**R**	**L**	**R**
PCC	1.0	1.0	1.0	.50	.67	.75	0.0	0.0	0.33	.50	.60	.67

Bold letters indicate correct choices, Italic and underlined letters indicate incorrect choices

PCC proportion correct at the level attempted, *L* left box, *R* right boxes

Procedure

To prevent satiation, dogs were fasted for four hours preceding the test. Before training, each dog was allowed to freely explore the testing and waiting areas. This acclimation period concluded when the dog ceased all exploratory behaviors and at least five minutes had elapsed.

In both training and testing, trials consisted of a demonstration by E1 and a choice by the dog. During demonstrations, E1 baited one of the boxes while the dog observed from the holding area. Starting from the far end of the testing area, E1 first attracted the dog's attention by taking one step forward while simultaneously holding up a treat and calling the dog's name. Keeping the treat raised, E1 walked towards, and then placed a treat on, one of the boxes. If the dog looked away during this demonstration, E1 paused and called to the dog until the dog redirected its attention towards E1 and the treat. After returning to the starting location, E1 turned to face the holding area and gazed at the top of the doorway. The choice portion of each trial began when E2 opened the gate while simultaneously speaking the specific release word used by the dog's owner. The dog was then given up to one minute to retrieve the treat but was only allowed to visually inspect a single box.

In between each trial, E1 returned the dog to the holding area and remained by the gate until E2 placed a foam board in front of the gate to block the dog's view of the testing area. This allowed E1 to remove an un-retrieved treat from the previous trial while maintaining the association that dogs were required to learn for this task (i.e., the box approached and touched by E1 always indicated the location of hidden food). After E1 collected a new treat and returned to the starting location, E2 indicated to E1 the next box to be baited in the upcoming trial.

Before testing, each dog completed three stages of training in which the choice portion of the trial immediately followed the demonstration. In Training Level 1 (T1), only one box was present in each trial and the treat remained visible throughout the demonstration and choice (i.e., the treat remained in view after being placed on the box platform in the demonstration). In Training Level 2 (T2), only one box was present in each trial but the occluder was added to the box shield so that the treat was not visible once placed on the box. In Training Level 3 (T3) and throughout testing, both boxes were present for each trial and treats were not visible once placed on a box.

With the box occluders in place, the dog was required to walk past the sides of the box shields to see which of the boxes held the treat. Thus, in testing, the first box visually inspected by the dog after entering this area of visibility was scored as the dog's choice for that trial. Choices were coded as a correct if the inspected box corresponded to the baited box containing the treat. Choices were coded as incorrect if the inspected box corresponded to the un-baited box or if the dog inspected neither box within one minute. Choices were determined by E1 and later confirmed from videos recorded by two synchronized cameras. The north camera was used to determine when treats became visible (Fig. 1a) and the south camera was used to track the dog's eyes (Fig. 1b).

Given that only one box was present in T1 and T2, the dog was not prevented from exploring the testing area before retrieving the treat in these training levels. However, a choice was only coded as correct if the dog walked directly to the correct box.

Experiment I protocol

Training trials were presented in blocks of two and training blocks were repeated following an incorrect choice on either trial (a failed block). Correct choices on both trials in a block (a completed block) lead to a "step up" to the next level, but only if the proportion of correct choices (PCC) across all trials at the current training level was also greater than 50% (Table 1: after trials 2, 6, and 12, but not after trial 10). If the dog failed three blocks at the same training level, the test was ended.

During testing, a delay varying from 15 to 180 s was implemented between the demonstration and choice, and the baited box followed a pseudorandomized order (see Fig. 2 for example of test progression). Trials at each delay were repeated in blocks of two and the length of the delay
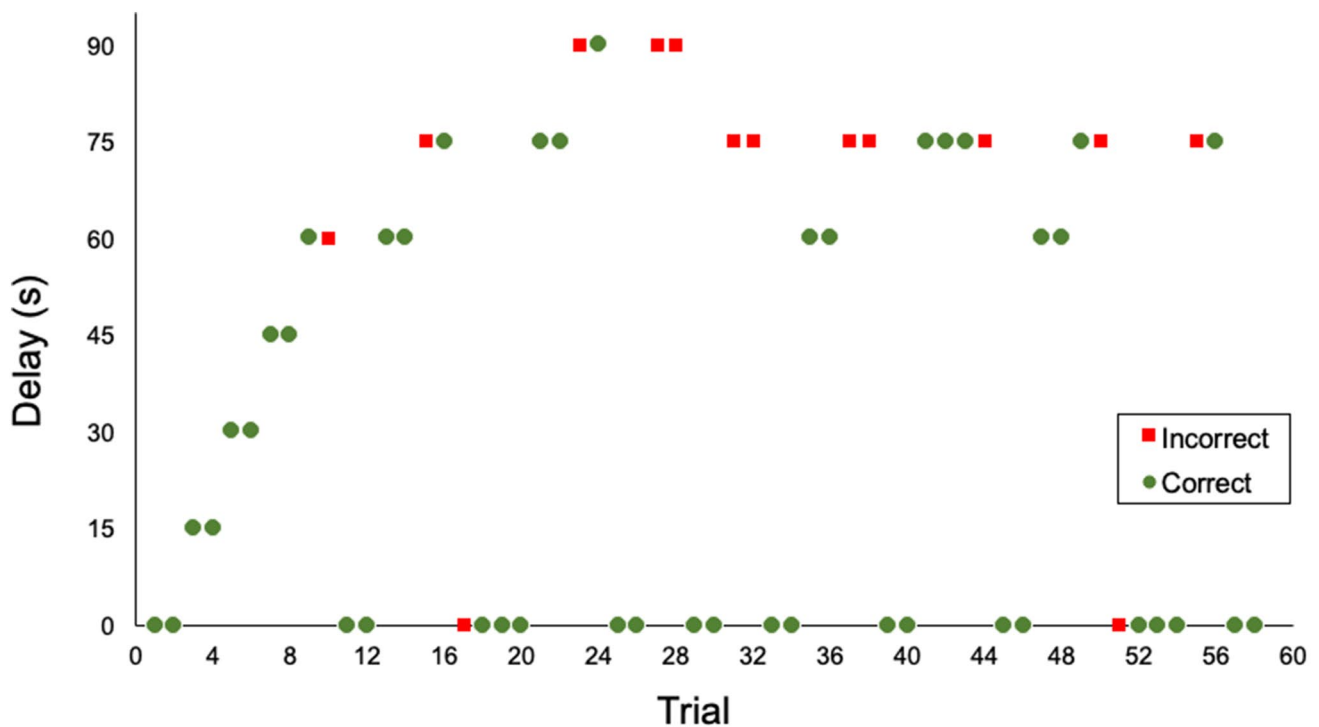
Fig. 2 Example of test progression in Experiment I. Green circles indicate correct choices. Red squares indicate incorrect choices

increased or decreased from block to block in increments of 15 s. A completed block resulted in a step up to a longer delay, but only if PCC at the given delay was also greater than 50%. If PCC at the delay was equal to or less than 50% following a completed block (e.g., Fig. 2: trials 41–42), the delay was repeated (e.g., Fig. 2: trials 43–44).

Following a failed delay block, the dog was given a block of two trials in which a delay was not implemented between the demonstration and choice (i.e., "refreshers"). When the dog failed a refresher block (e.g., Fig. 2: trials 17–18), a second refresher block was implemented (e.g., Fig. 2: trials 19–20). If the dog failed the second refresher block, the test ended. Following a successfully completed refresher block, the dog returned to the delay staircase.

The first time the dog returned to the delay staircase (e.g., Fig. 2: trials 27–28), the previously attempted delay (e.g., Fig. 2: trials 23–24) and box were repeated. If the dog failed this block of delay trials (e.g., Fig. 2: trials 27–28) and then completed another block of refresher trials (e.g., Fig. 2: trials 29–30), the delay was decreased following the second (e.g., Fig. 2: trials 31–32) and third (e.g., Fig. 2: trials 35–36) consecutive returns to the delay staircase. If the dog failed the delay block after the third consecutive return, the test ended (e.g., if the dog made an incorrect choice in Fig. 2: trial 35 or 36). In addition, the test ended if the dog failed three consecutive blocks at the same delay (e.g., Fig. 2: after trials 55–56). If none of the above criteria for ending the

test were met, the session continued for up to one hour. At the end of the test, two refresher trials were implemented to confirm that the dog was not satiated and to assess whether the dog continued to choose the box most recently visited by E1 throughout the test (e.g., Fig. 2: trials 57–58).

This up–down staircase procedure did not require stringent progression criteria to produce reliable evidence that the dog's memory span was at least as long as a completed delay. Given that time is additive and that the spatial-matching rule was the same in T3, refreshers, and all delays, correct choices in longer delays also functioned as correct choices in shorter delays. For example, the dog in Fig. 2 made correct choices in 13 of 14 trials with delays of 15–60 s and, therefore, provided highly reliable evidence of a 15 s memory span despite attempting only two trials at the 15 s delay. The same reasoning holds for delays of at least 30 s and delays of at least 45 s. Thus, additional trials at these shorter delays would have been unnecessary and inefficient.

The up–down procedure and greater difficulty of remembering the correct location at longer delays also mitigated potentially detrimental consequences of allowing the dog to progress beyond its true SWM threshold. As demonstrated in trials 21–22 of Fig. 2, a dog with a 60 s memory span was not unlikely to progress beyond the 75 s delay by randomly guessing the correct box in two consecutive trials ($p = 0.25$ if the delay was novel and $p \leq 0.19$ if the dog had

Table 2 Example of training progression in experiment II

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Level	1	1	2	2	2	2	3	3	3	2	2	3	3	3	3
Choice	**L**	**R**	**L**	*R*	**r**	**L**	*R*	*r*	*r*	**L**	**R**	**L**	**R**	**R**	**L**
PCC	1.0	1.0	1.0	.50	.67	.75	0.0	0.0	0.0	.80	.83	.25	.40	.50	.57

Bold letters indicate correct choices, italic and underlined letters indicate incorrect choices, bold italic lowercase letters indicate correction trials

PCC proportion correct at the level attempted, *L* left box, *R* right boxes
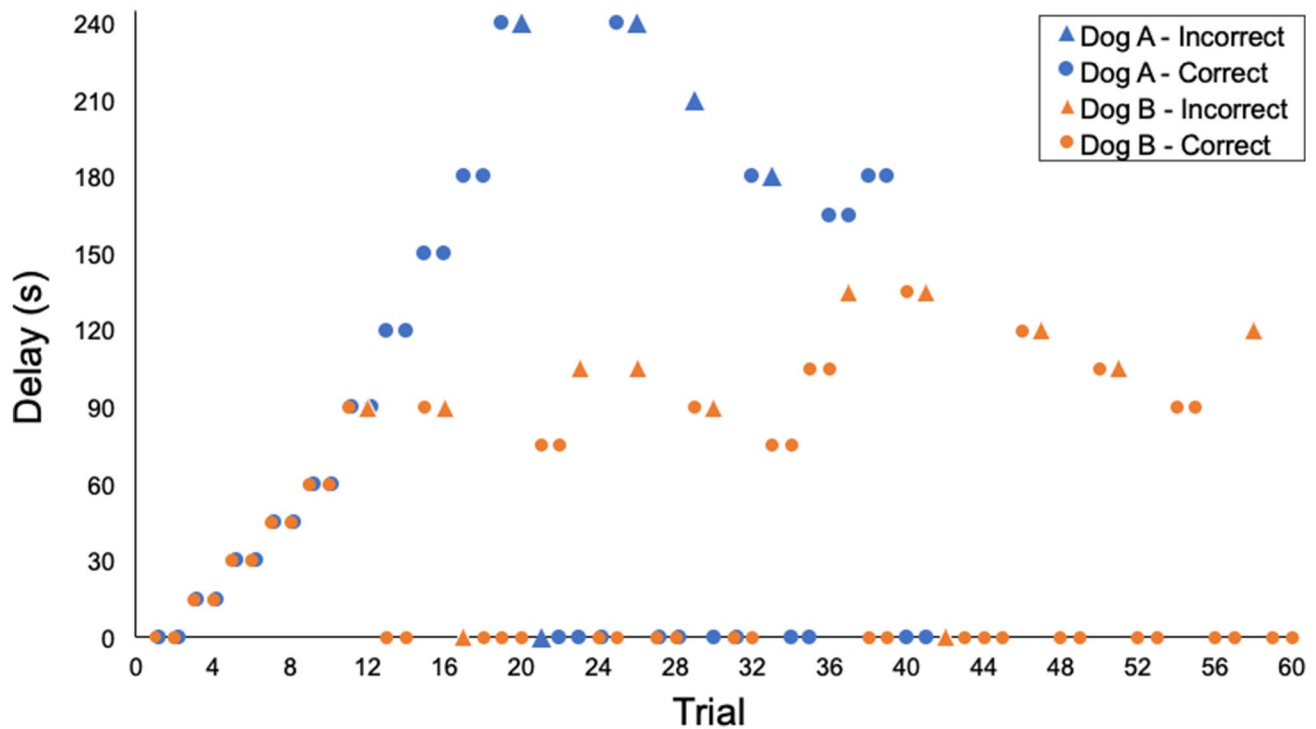

Fig. 3 Examples of test progression in Experiment II. Blue and orange points represent example dogs A and B, respectively. Dots indicate correct choices, triangles indicate incorrect choices. Note that dog A attempted longer delays and, therefore, fewer trials within the hour-long test (colour figure online)

previously attempted the delay). However, the probability of completing consecutive blocks by random chance was low ($p = 0.06$ if both delays were novel and $p \leq 0.03$ if the dog had previously attempted blocks at these delays). Additionally, the dog was unlikely to avoid repeating a delay that was previously completed by chance. Indeed, the example dog in Fig. 2 was ultimately required to complete six more blocks at the 75 s delay after erroneously progressing to the 90 s delay. Most importantly, the test provided evidence that the dog's memory span was at least 60 s (7 of 8 correct) but less than 75 s (8 of 16 correct).

Experiment II protocol

The protocol for Experiment 2 differed from Experiment 1 in the following respects. Both in training and testing, trials were not necessarily implemented in blocks of two. In

training, the baited box alternated after a correct choice. An incorrect choice immediately resulted in a correction trial in which the training level and baited box were repeated. Following a correct choice on a correction trial, the training step was repeated at the other box (Table 2: trials 4–6). If the dog chose incorrectly on a correction trial, that trial was repeated. Two successive incorrect choices on correction trials resulted in a step down to a correction trial on the previous training level (Table 2: trials 7–10). If at any point a dog did not retrieve food in four successive trials, the session was ended.

During the test, a delay of 15–300 s was implemented between the demonstration and choice. From 0 to 59 s, delays increased by 15 s; from 60 to 179 s, delays increased by 30 s; and from 180 s on, delays increased by 60 s (Fig. 3: dog A, trials 1–20). An incorrect choice on any delay trial immediately resulted in a zero-delay refresher trial (e.g.,

Table 3 Experiment II analyses

Purpose	Analysis	Criterion	Predictors	n	Tables/Figures
Task acquisition and retention	Binomial tests	PCC in zero-delay trials[a] (one test per dog)		6–33[b]	Figure 6 Tab. S4
Task acquisition and retention	Binomial tests	PCC in delays trials[a] (one test per dog)		8–29[b]	Figure 6 Tab. S4
Age-Dependent Performance	Quadratic regression	CMS	$AGE + AGE^2$	23	Figure 7
Delay-Dependent Performance[a]	Binomial GLMM	Trial outcome	$AGE + AGE^2 + Delay + Trial + AGE*Delay + AGE^2*Delay + AGE*Trial + AGE^2*Trial + (1 + Delay + Trial \mid Dog)$	696	Table 4
Head and body orientations	Binomial GLMM	Delay trial outcome	$H.C.Dl + B.C.Dl + H.F.Dl + B.F.Dl + H.F.Dm + B.F.Dm + (1 \mid Dog)$	426	Table 5 Table 6
Perseveration	Quadratic regression	|Side bias|	$AGE + AGE^2$	23	Figure 8
Delay optimization	5 × quadratic regressions	PCC in the 15 s, 30 s, 45 s, 60 s, and 90 s delays	$AGE + AGE^2$	20–22[c]	Table 7 Figure 9

Predictors in the final GLMM of trial outcome are in bold text. GLMM equation notation follows from the package lme4 in R. For head and body orientations

GLMM generalized linear mixed model (logit error distribution), *PCC* proportion or correct choices, *CMS* cumulative memory score, *H* head, *B* body, *C* oriented towards the correct box, *F* oriented forwards (towards the testing room), *Dl* during the entire delay, *Dm* during the entire demonstration

[a] Binomial tests compared observed to expected number of correct choices in a given number of attempts

[b] Number of trials differed among dogs

[c] Number of dogs differed among delays

Fig. 3: dog B, trial 23). After one correct choice on a refresher trial, the dog was given a second refresher trial on the other box. Following an incorrect choice on a refresher trial, the dog was given a correction trial which consisted of another refresher trial on the same box. Correction trials were repeated until a correct choice was made, after which another refresher at the other box was implemented. The dog returned to the delay staircase following consecutive correct choices if PCC in refresher trials was above 50%.

The first time a dog returned to the delay staircase, the previously attempted delay and box were repeated. However, in successive returns, the previously attempted box was repeated but the delay decreased. If the failed delay was between 0 and 180 s, the delay decreased by 15 s (Fig. 3: dog B, trials 40–55). If the failed delay was more than 180 s or, the delay decreased by 30 s (Fig. 3: dog A, trials 25–33).

After failing the first block of 15 s delay trials, one dog (Greta) developed a strong side bias and failed to return to the delay staircase despite ten refresher and correction trials. Therefore, she was given a counter-training procedure which required four consecutive correct choices in both T2 and T3.

Analysis overview

Due to a low rate of test completion in Experiment I, we used only descriptive statistics and graphical analyses of test outcomes to inform the design of the second assessment. Statistical analyses for Experiment II are outlined in Table 3. Tests of assumptions and case-wise diagnostics for these analyses are described in supplement S2. All data were analyzed in R version 3.4.1. Generalized linear mixed models (GLMMs) were constructed and tested using the package "lme4" (Bates et al. 2015).

Body size and sex effects

The prevalence of CDS may be similar in breeds which differ greatly in size (Salvin et al. 2010, 2012). In addition, Watowich et al. (2020) recently found that lifespan changes in cognitive performance did not depend on the average lifespan of a dog's breed. However, lifespan is inversely related to body size in dogs and physiological deterioration may progress more rapidly in large breeds (Kraus et al. 2013). Thus, to account for possible effects of body size, all statistical analyses were repeated with age measured in months (AGE) and as a proportion of the dog's size-predicted lifespan (PSL). Size-predicted lifespan was calculated using the following equation from Greer et al. (2007): predicted lifespan (years) = 13.62 + (0.0702 × height (inches))—(0.0538 × weight (pounds)). The relative strength and significance of predictors did not change in any analysis when age was treated as PSL. Therefore, only analyses of AGE are reported.
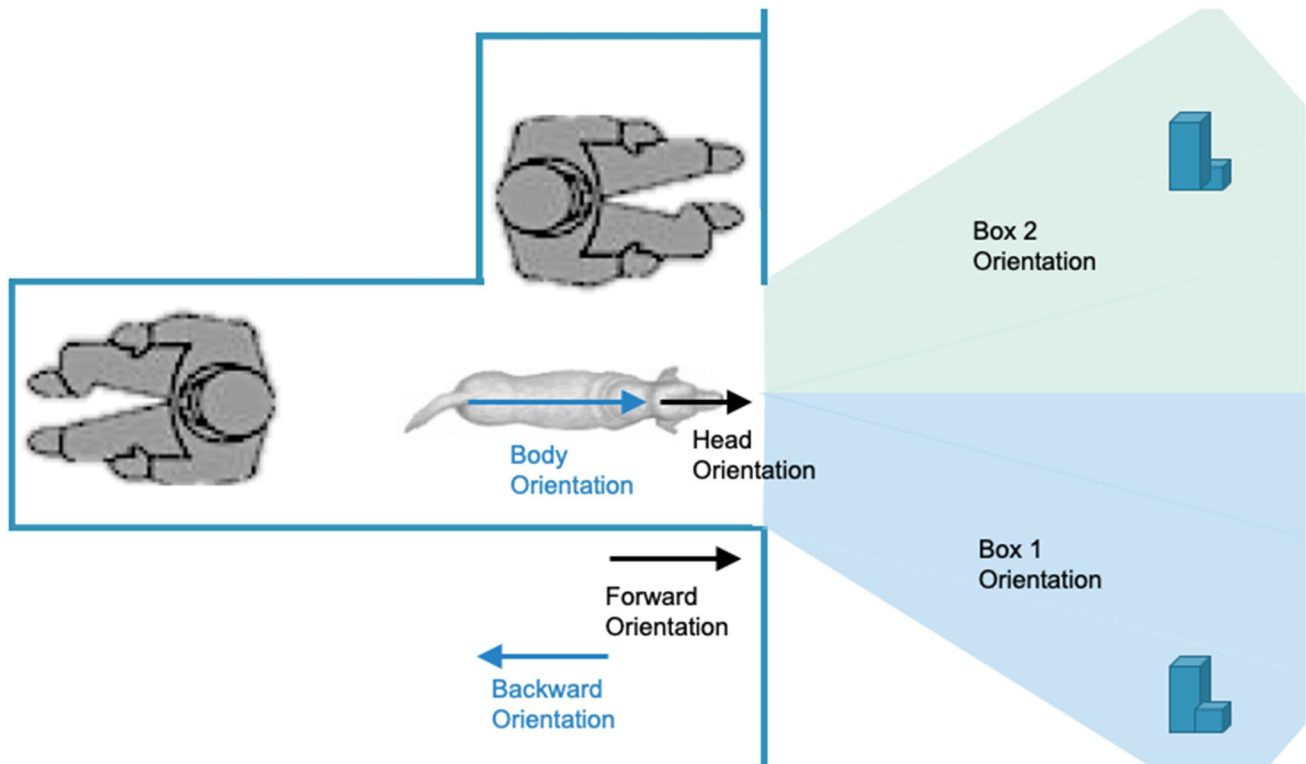
Fig. 4 Diagram of the orientation coding criteria

Sex may affect spatial learning and memory in cognitive aging tests for dogs (e.g., Mongillo et al. 2017). Thus, all analyses were first conducted with sex included as a covariate (see supplement S3). Sex did not predict performance in any analysis and lead to issues of non-convergence in analyses which used GLMMs. In addition, accounting for sex did not lead to any meaningful change in the outcome of an analysis.

Experiment II measurements

Proportion of correct choices

The validity of this assessment was contingent on evidence that each dog attempted to search for food at the box visited by E1 during the most recent demonstration (i.e., the correct box). Moreover, incorrect choices in delay trials could not be attributed to SWM deficits if the dog failed to follow this spatial-matching rule. Thus, for each dog, PCC in zero-delay trials (the third step of training and refreshers) and PCC in delay trials were used to examine whether the dog learned the spatial-matching rule and chose accordingly throughout the test.

Cumulative memory score

To measure overall performance, the sum of the delays after which the dog chose the correct box was used as a weighted Cumulative Memory Score (CMS). Only the first 18 delay trials (the fewest completed by any dog) were included in this score. Thus, CMS measured the total amount of time in the first 18 delay trials that the dog remembered the location of the treat.

Orientation behaviors

To investigate whether the dog used non-mnemonic orientation strategies to locate the treat, a coder recorded from video whether the dog's head and body (coded separately) remained oriented towards the correct box for the entirety of the delay. Orientation was scored as correct when the long axis of the head or body pointed towards the correct box, between the midpoint and upright of the doorway (Fig. 4). To measure initial and sustained attention, the coder scored whether the dog's head and body (coded separately) remained oriented toward the testing area for the entirety of the demonstration and the delay (coded separately). Orientation was scored as forward when the long axis of the head or
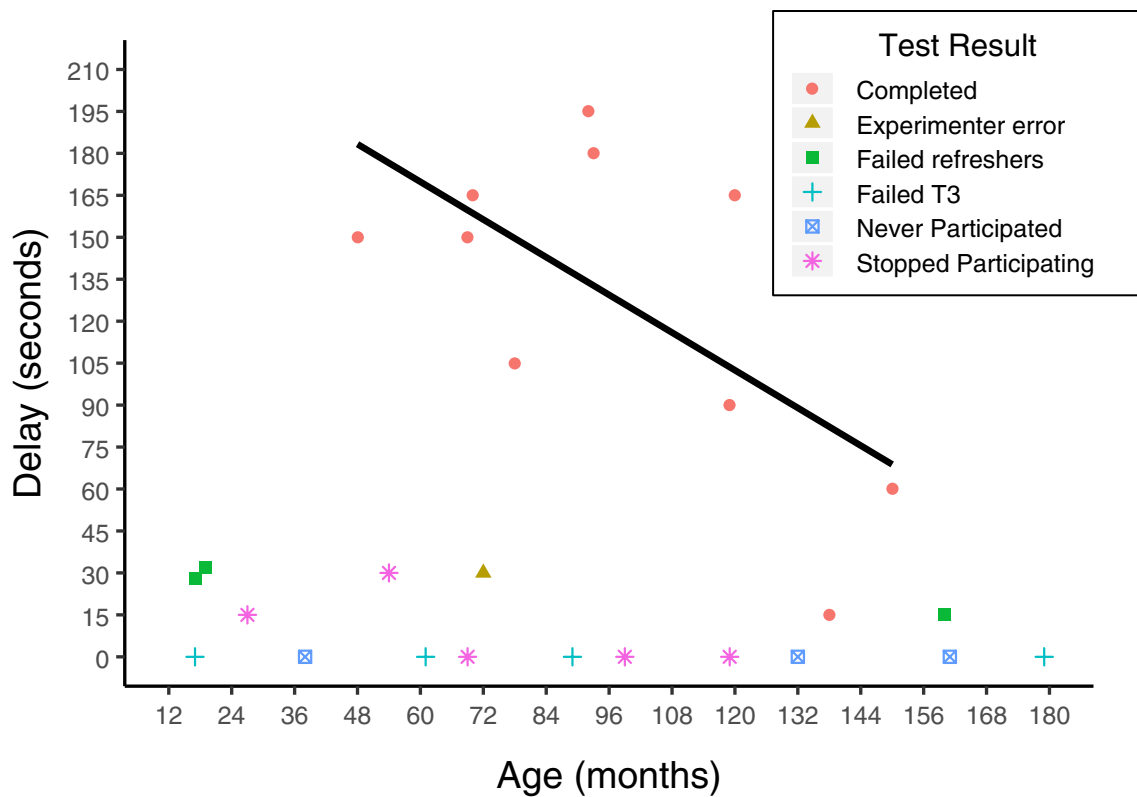
Fig. 5 Longest delay attempted in Experiment I by age. The regression line is fitted only to data of dogs that completed the test (red dots). Blue squares indicate dogs that never participated. Pink strikes indicate dogs that stopped participating. Blue crosses indicate dogs that participated fully but failed to pass training. Green squares indicate dogs that passed training but failed two consecutive blocks of refresher trials. One test was ended prematurely due to an experimenter error (green triangle) (colour figure online)

body pointed towards the testing area between the uprights of the doorway (Fig. 4). To assess interrater reliability, 25% of the videos were scored by a second coder. For each orientation behavior, coders agreed on at least 98% of trials. Due to video inadequacies, orientations could not be coded for 11 (2.5%) trials.

Box preference (side bias)

Perseveration rather than inadequate memory span may result in incorrect choices. The dog may incorrectly choose its preferred box despite correctly recalling that E1 visited the other box in the most recent demonstration. In the first experiment, the dog completed the same number of trials at each box. Therefore, the number of errors the dog made at each box was used to assess whether it was hindered by a perseverative tendency to choose its preferred box. In the second experiment, correction trials were implemented to combat the development of a side bias. Thus, dogs were not required to complete the same number of trials at each box. As a result, the magnitude of the dog's box preference was used to measure his perseverative tendency. Specifically,

side bias was calculated as the absolute value of the difference between PCC on Box 1 and PCC on Box 2.

Results

Experiment I

The completion rate in the first experiment was too low to test whether age significantly predicted performance. Thus, to determine which procedural modifications were needed to allow dogs of all ages to complete the test, the longest delay attempted (i.e., the highest step reached by the dog) and the reason for ending the test were plotted as a function of age (Fig. 5). To assess whether such modifications were likely to produce a viable test for ARCD, a regression line was fitted to the data of dogs that completed the test.

Ten of the 26 dogs did not pass training. Three of these dogs stopped participating, three never participated, and four made incorrect choices in three consecutive blocks at T3. Two dogs stopped participating after the first incorrect choice in a delay trial and three made incorrect choices in consecutive blocks of refresher trials. One test was ended
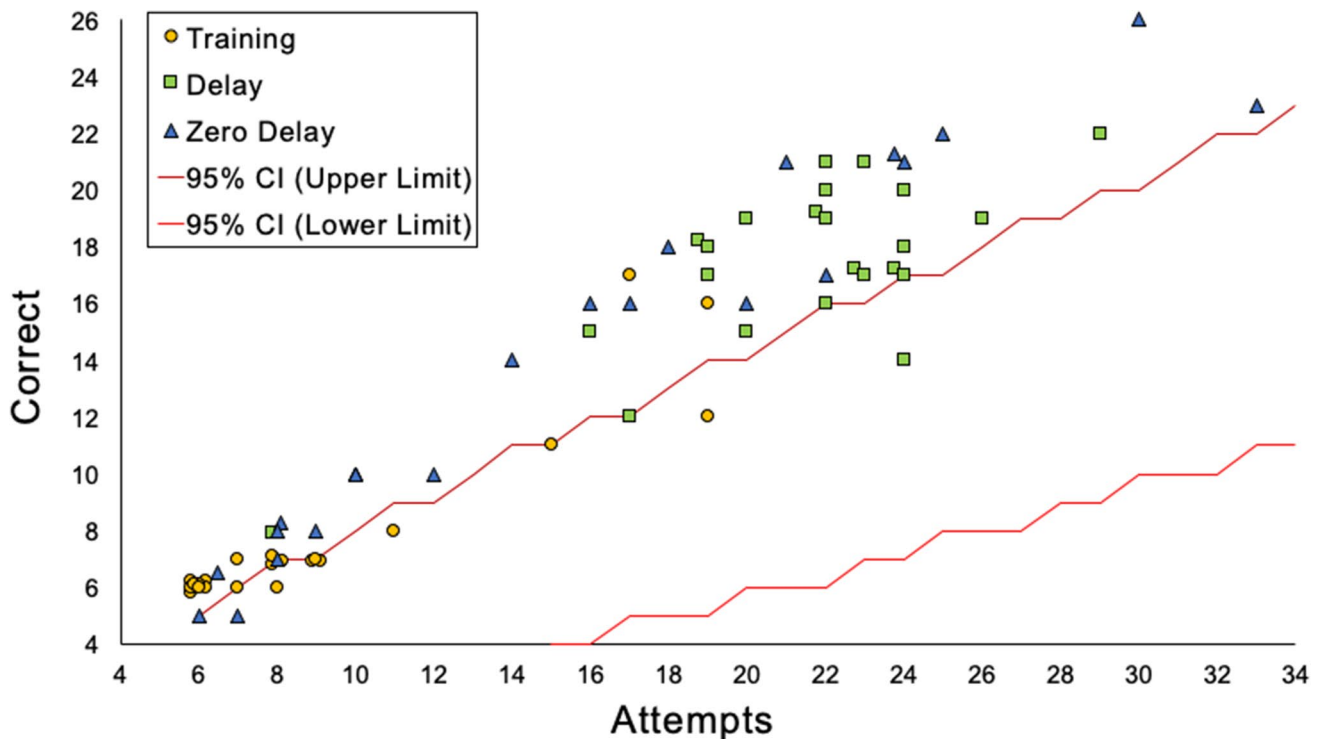
Fig. 6 Correct choices by trials attempted in Experiment II. Each dog is represented by three points. The red lines indicate the maximum and minimum number of correct choices within the 95% confidence interval for random chance. Points within these lines did not significantly differ from chance. Data for dogs that did not differ from chance in either zero-delay trials (green squares) or delay trials (blue triangles) are provided in Table 4. Dogs were not required to exceed chance in training trials (yellow dots) because T3 and refresher trials were identical assessments for task comprehension (colour figure online)

prematurely due to an experimenter error. Of the seven dogs that participated fully but failed to complete the test, five provided evidence of a strong side bias; each of these dogs made at least four times as many errors when the preferred box was not baited. Collectively, these dogs made 28 errors when the non-preferred box was baited and three errors when the preferred box was baited.

Among the ten dogs that completed the test (AGE: $\bar{x} \pm s = 97.7 \pm 33.0$ months), the longest delay attempted decreased rapidly with age (delay = 138.3—1.13 * centered AGE). However, no dogs younger than 4 years and older than 12 years completed the test.

Experiment II—assessment validity

Test completion

Of the 25 dogs that returned, 23 completed the test. The two dogs that did not complete the test never participated in either experiment. These dogs were among the oldest tested (AGE = 163 and 134 months). The other dog that did not participate in Experiment I acquired and retained the spatial-matching rule in Experiment II and completed this modified assessment but still did not make a choice in 25% of all trials

(AGE = 40 months). This dog was only included in analyses of PCC and side bias, which did not include no-choice trials. Of the remaining 22 dogs, 18 participated in all trials and four participated in at least 90% of trials.

Proportion of correct choices

For each dog that participated, two binomial tests were conducted to assess whether PCC in zero-delay and delay trials differed from chance (E(PCC) = 0.5). PCC was significantly above chance in zero-delay trials for 20 of the 23 dogs (Fig. 6). The other three dogs were each far above chance in delay trials ($p < 0.0001$; Table S4). PCC was significantly above chance in delay trials for 19 of the 23 dogs (Fig. 6). The other four dogs were each far above chance in zero-delay trials ($p < 0.0005$; Table S4).

Cumulative memory score

A regression analysis of CMS by centered AGE (AGE. CENT) was conducted to assess whether age predicted cumulative performance. However, an inverted U-shaped relationship was observed between AGE.CENT and CMS. Thus, a coefficient for the predictor AGE.CENT^2 was also
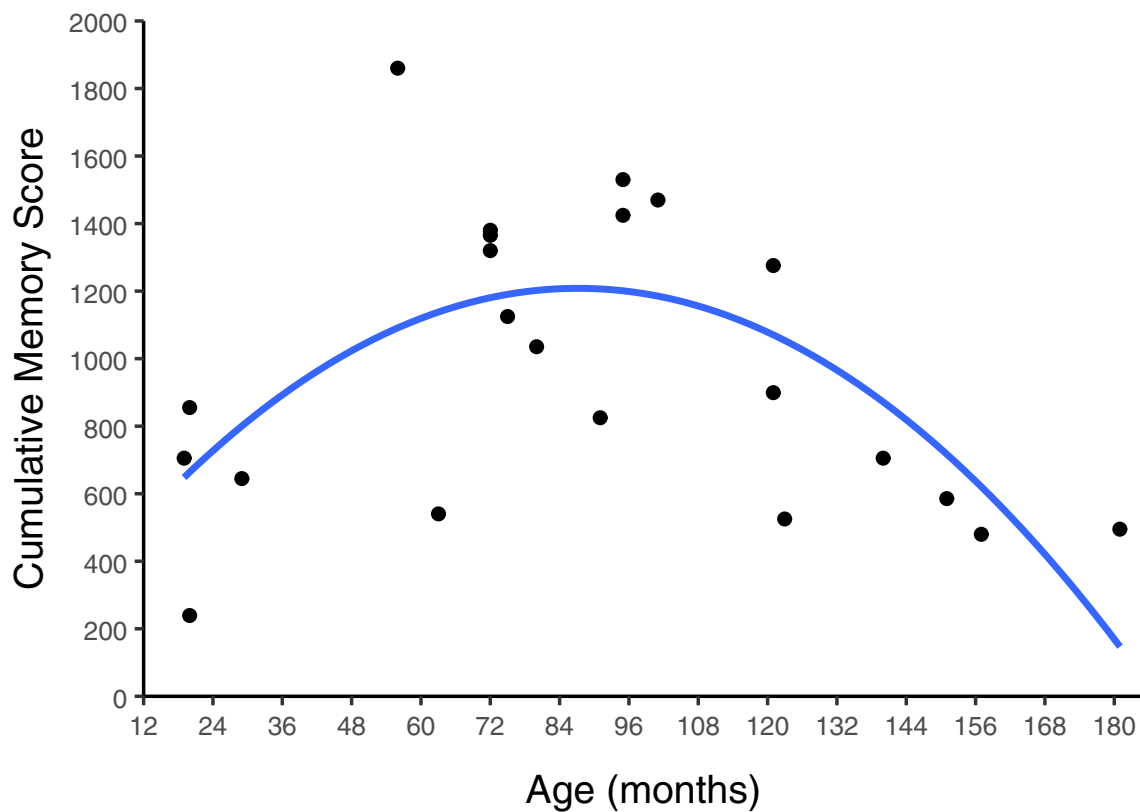
Fig. 7 Cumulative memory score (CMS) in Experiment II by age. Each black dot corresponds to one dog's observed CMS—the total amount of time in the first 18 delay trials that the dog remembered the correct location of the hidden treat. The equation for the blue OLS regression line is, $CMS = 292.31 + 21.02*AGE - 0.12*AGE^2$, where *AGE* age in months

included in the regression equation. An F test was used to test the overall prediction of the model and two-tailed t tests were used to test the significance of the predictors. AGE. CENT and $AGE.CENT^2$ yielded significant overall prediction of CMS, $F(2,19) = 7.24$, $p = 0.005$; $r^2 = 0.43$. The linear effect of AGE.CENT was not significant, $b_1 = -0.40$, $SE_{b1} = 1.65$; $t(19) = -0.24$, $p = 0.81$, but $AGE.CENT^2$ significantly predicted CMS, $b_2 = -0.12$, $SE_{b2} = 0.03$; $t(19) = -3.73$, $p = 0.001$. Memory scores ($\bar{y} = 967.5$, $SE_y = 92.41$) were higher in middle-aged dogs than in young and old dogs (Fig. 7).

Test–trial outcome

To assess whether performance during the test (i.e., in delay and refresher trials) was age- and delay-dependent, a binomial regression analysis of test–trial outcome (correct or incorrect choice) was conducted using a GLMM with a logit-link function (Table 4). A hypothesis-driven (full) model was first constructed with random subject slopes for delay and trial, random intercepts for subjects, and fixed effects for delay, trial number, AGE, AGE^2, delay*AGE, delay*AGE^2, trial*AGE, and trial*AGE^2. The overall fit of

this model was then compared to that of the maximal converging (final) model (tolerance = 0.001) using Akaike and Bayesian Information Criteria (AIC and BIC). To aid in convergence, continuous predictors were rescaled to Z scores. Profiled confidence intervals of fixed-effect estimates and of random-effect variances were calculated using parametric bootstraps. In addition, likelihood ratio tests of competing models ($\alpha = 0.05$) were used to test the significance of interactions and random effects (this was not possible for the main effects).

The full model did not converge due in part to overfitting with interaction terms and random slopes. Furthermore, the random slopes for trial and delay were perfectly correlated with the random intercepts. The final model included only the main effects and random intercepts. Removing the interactions and random slopes improved overall fit (decreased AIC and BIC) but did not change whether the profiled confidence intervals of the remaining predictors overlapped with zero (i.e., the same predictors were significant). Both models are summarized in Table 4.

In the final model, the probability of choosing the correct box decreased significantly with delay length, $X^2(1) = 20.89$, $p < 0.0001$, but not trial number, $X^2(1) = 1.29$, $p = 0.26$. The

Table 4 Delay sensitivity—GLMM of test-trial outcome

Model		Estimates					Random effects			Model fit		
Predictor	Type	ß	SE	ß CI	X^2	p	s^2 CI	X^2	p	AIC	BIC	r^2 r(df)
Full										614	682	0.96
AGE	F	2.46	0.13	[1.64, 3.32]	–	–	–	–	–			
AGE^2	F	− 2.25	0.51	[− 3.08, − 1.38]	–	–	–	–	–			
Delay	F, RS	− 0.61	0.14	[− 0.87, − 0.34]	–	–	[.01, .43]	0.17	0.98			
Delay*AGE	F	0.81	0.54	[− 0.02, 1.79]	2.23	0.69	–	–	–			
Delay*AGE^2	F	− 0.82	0.54	[− 1.80, 0.01]	2.32	0.68	–	–	–			
Trial	F,RS	− 0.10	0.13	[− 0.35, 0.20]	–	–	[.01, .33]	0.99	0.09			
Trial*AGE	F	− 1.57	0.47	[− 2.51, − 0.75]	9.34	0.05	–	–	–			
Trial*AGE^2	F	1.55	0.47	[0.77, 2.55]	8.92	0.06	–	–	–			
Intercept	I	1.61	0.51	[1.43, 1.93]	–	–	–	–	–			
Subject	RI	–	–	–	–	–	[.01, .40]	0.09	0.99			
Final										608	635	0.92
AGE	F	2.78	0.87	[1.45, 4.41]	11.68	.0006	–	–	–			
AGE^2	F	− 2.58	0.86	[− 4.20, − 1.30]	10.54	.001	–	–	–			
Delay	F	− 0.66	0.16	[− 0.94, − 0.44]	20.89	< .0001	–	–	–			
Trial	F	0.20	0.17	[− 0.04, 0.52]	1.29	.26	–	–	–			
Intercept	F	1.86	0.20	[1.50, 2.26]	–	–	–	–	–			
Subject	RI	–	–	–	–	–	[0, 0.82]	0.48	0.50			

Chi-square and p values were obtained from likelihood ratio tests of the difference in total prediction between the full model and the nested model without the predictor. Profiled confidence intervals for ß were calculated using parametric bootstraps

F fixed effect, RS random slope, RI random intercept, β estimate (predictor coefficient), s^2 95% CI confidence interval of the random effect variance (values near zero are not significant), AIC and BIC akaike and bayesian information criterion (smaller values indicate better fit), r^2/r(df) sum of squared Pearson residuals divided by residuals degrees of freedom (deviations from 1.0 larger than |0.2| may indicate over/underdispersion).

effects of AGE, $X^2(1) = 11.68$, $p = 0.0006$, and AGE^2 were significant, $X^2(1) = 10.54$, $p = 0.001$. Middle-aged dogs chose correctly more often than young and old dogs. The random intercept for study subject was not significant, $X^2(1) = 2.25$, $p = 0.13$.

Experiment II—exploratory analyses

Head and body orientations

Because they did not apply to refresher trials, delay orientations could not be treated as covariates in the previously described GLMMs, which estimated the outcome of either a refresher or a delay trial. Therefore, separate exploratory analyses were used to examine whether head and body orientations predicted the outcome of a delay trial. However, the previous analysis indicated that this reduced set of data could not support a model with more than a limited set of predictors (i.e., a model with coefficients for all six orientation behaviors, AGE, AGE^2, delay, trial, the interactions of these fixed effects, and random effects would be overfit). Thus, the predictive value (i.e., utility) of each orientation behavior was first estimated using a GLMM of trial outcome

with a binomial (logit) error distribution, fixed effects for each orientation (orientation maintained or broken) and random intercepts for study subjects (see Table 5).

Only body orientated forwards during the delay (B.F.Dl) significantly predicted trial outcome, $X^2(1) = 8.46$, $p = 0.004$. Therefore, to explore whether B.F.Dl and, in turn, sustained engagement might account for the age effects observed in the primary analyses, another GLMM of delay-trial outcome (logit error distribution) was constructed with fixed effects for B.F.Dl, AGE, AGE^2, delay, B.F.Dl*AGE, and B.F.Dl*AGE^2. Subject intercepts and slopes for B.F.Dl were included as random effects. The final model is outlined in Table 6.[1] Confidence intervals for the conditional effects

[1] Trial number was not included as a predictor in this model given that the omission of refresher trials created highly discontinuous and individualized trial number data (i.e., numerous singularities). In addition, trial number yielded no prediction of trial outcome in the previous models. Likewise, random subject slopes for delay, and coefficients for the interactions between delay and age were also omitted because these effects lead to singularities and non-convergence but did not predict trial outcome, improve model fit, or change the prediction of other effects in the previous models.

Table 5 Predictive value of orientation behaviors—GLMM of delay trial outcome

Dimension	Predictor	Levels	Group 1 (no) n	Group 2 (yes) PCC $(M \pm SE)$	LRT n	Dimension PCC $(M \pm SE)$	Predictor X^2	Levels p
Rehearsal	Head correct (delay)	Correct orientation (yes/no)	454	0.78 ± 0.02	20	0.70 ± 0.11	1.11	0.29
	Body correct (delay)		362	0.75 ± 0.02	112	0.85 ± 0.03	2.45	0.12
Engagement	Head forward (delay)	Forward orientation (yes/no)	380	0.77 ± 0.02	94	0.77 ± 0.04	0.13	0.72
	Body forward (delay)		113	0.65 ± 0.04	361	0.81 ± 0.02	8.46	0.004
	Head forward (demonstration)		137	0.79 ± 0.04	337	0.77 ± 0.02	1.23	0.27
	Body forward (demonstration)		74	0.74 ± 0.05	400	0.78 ± 0.02	0.29	0.59
Individual variation	Subject	Random intercepts	–	–	–	–	1.34	0.25

Group 1: trials in which the dog did not maintain the corresponding orientation. Group 2: trials in which the dog did maintain the corresponding orientation. The full model included all orientation behaviors and the random intercept. See Table 3 for remaining abbreviations

LRT likelihood ratio test of nested models, *n* number of trials in which the dog did or did not maintain the orientation

Table 6 Engagement, delay, and age—GLMM of delay-trial outcome

Model Predictor	Type	Estimates ß	SE	ß CI	X^2	p	Random effects s^2 CI	X^2	p	Model fit AIC	BIC	r^2 $r(df)$
Full										502	543	0.95
Age	F	1.86	0.78	[0.45,3.35]	–	–	–	–	–			
Age^2	F	− 1.72	0.72	[− 3.09,-0.46]	–	–	–	–	–			
B.F.Dl	F,RS	0.76	0.28	[0.11,1.26]	–	–	[0.0,0.84]	0.92	0.63			
B.F.Dl*AGE	F	− 0.08	0.96	[− 2.09,1.61]	0.01	0.94	–	–	–			
B.F.Dl*AGE^2	F	− 0.14	0.92	[− 1.78,1.72]	0.02	0.88	–	–	–			
Delay	F	− 0.26	0.14	[− 0.53,0.002]	3.49	0.06	–	–	–			
Intercept	F	0.78	0.22	[0.35,1.31]	–	–	–					
Subject	RI	–	–	–	–	–	[0.0,0.73]	0.19	0.91			

See Tables 3 and 5 for abbreviations

of age and age^2 on delay-trial outcome when the dog did (B.F.Dl = 1) and did not (B.F.Dl = 0) maintain orientation were simulated using the package "interplot" (Solt et al. 2015), following the method of Brambor et al. (2006).

Dogs that remained oriented towards the testing room during the delay were more likely to choose the correct box, 95% CI of β [0.11, 1.26]. The interactions between B.F.Dl and AGE, $X^2(1) = 0.01$, $p = 0.94$, and between B.F.Dl and AGE^2, $X^2(1) = 0.02$, $p = 0.88$, were not significant. Middle-aged dogs chose correctly more often than young and old dogs regardless of B.F.Dl. The conditional effect of AGE was significant in trials in which the dog remained oriented forwards, β = 1.77, 95% CI [0.46, 3.08], and in trials in which the dog turned away from the testing room, β = 1.86, 95% CI [0.38, 3.35]. Likewise, the conditional effect of AGE^2 was significant whether the dog did, β = − 1.84, 95% CI [− 3.18, − 0.52], or did not, β = − 1.71, [− 3.08, − 0.37], maintain forward body orientation during the

delay. The probability of choosing the correct box tended to decrease with delay length, but this trend was not significant, $X^2(1) = 3.49$, $p = 0.06$. Random subject intercepts, $X^2(1) = 0.19$, $p = 0.91$, and slopes for B.F.Dl, $X^2(1) = 0.92$, $p = 0.63$, were not significant.

Perseveration

To investigate whether inhibitory control may provide an alternative focus for future cognitive aging assessments using this paradigm, a regression analysis of side bias by AGE. CENT and AGE.CENT^2 was conducted. Trials in which the dog failed to make a choice were excluded from this analysis. AGE.CENT and AGE.CENT^2 yielded significant overall prediction of side bias, $F(2,20) = 5.37$, $p = 0.01$; $r = 0.35$. The coefficients for AGE.CENT, $b_1 = 0.0008$, $SE_{b1} = 0.0004$; $t(20) = − 2.15$, $p = 0.04$, and AGE.CENT^2 were also significant; $b_2 = 0.00002$, $SE_{b2} = 0.00007$; $t(20) = 2.87$, $p = 0.01$.
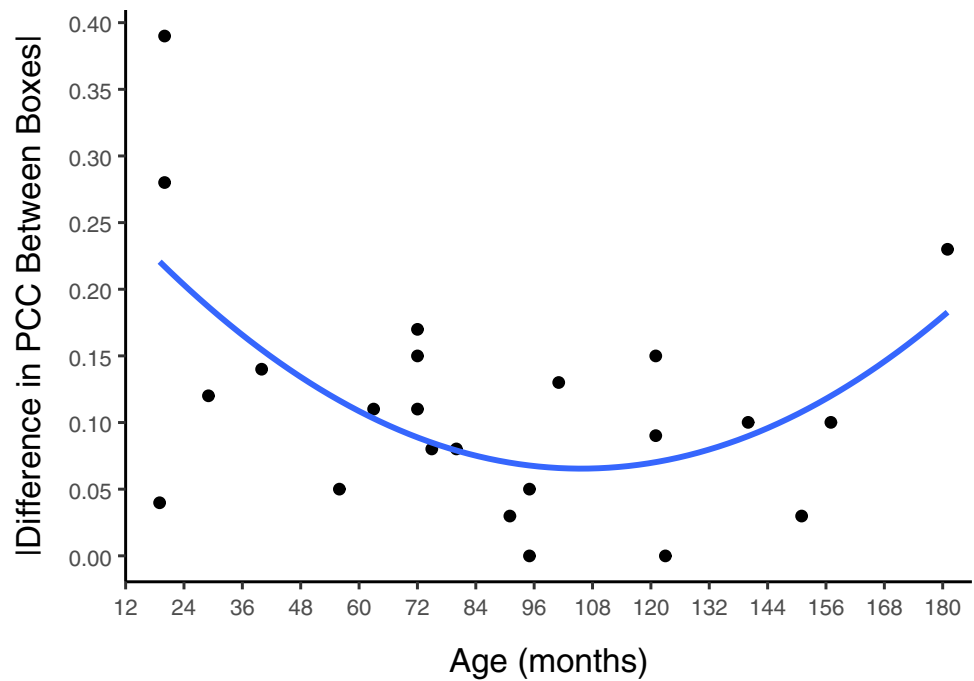
Table 7 Regression of age on PCC in each delay attempted by at least 20 dogs

Delay (s)	n	Model prediction			AGE.CENT		AGE.CENT²	
		r^2	F	p	T	P	t	p
15	22	0.20	2.35	0.12	2.12	0.05	− 2.16	0.04
30	22	0.15	1.63	0.22	1.80	0.09	− 1.77	0.09
45	22	0.33	4.67	0.02	0.33	0.75	− 3.06	0.01
60	22	0.35	5.11	0.02	− 1.31	0.21	− 2.71	0.01
90	20	0.28	3.38	0.06	0.90	0.38	− 2.57	0.02

Box preferences (mean = 0.11, $SE_y = 0.02$) were stronger in young and old dogs than in middle-aged dogs (Fig. 8).

Delay optimization

To identify the shortest delays that may be sensitive to ARCD in this paradigm, a separate regression analysis of PCC by AGE and AGE² was conducted for each delay that was attempted by at least 20 dogs. The *F* test and multiple r^2 of the model, and the two-tailed t tests of AGE and AGE² were then compared among delays (Table 7).

AGE and AGE² accounted for a significant proportion of variation in PCC at the 45 s ($r^2 = 0.33$, $p = 0.02$) and 60 s delay ($r^2 = 0.35$, $p = 0.02$) but not the 15 s ($r^2 = 0.20$, $p = 0.12$) or 30 s delay ($r^2 = 0.15$, $p = 0.22$). Two of eight old dogs (≤ 8 years) displayed deficits at the 15 s delay, compared to four old dogs at the 30 s and 45 s delays, and seven at the 60 s delay. The 60 s delay was also the longest delay reached by one old dog. Bivariate PCC by AGE plots for all delays are displayed in Fig. 9.

Discussion

Experiment I

Although the age-related decline in performance among dogs that completed the first test suggested that this assessment may be sensitive to ARCD, the low test-completion rate indicated that this version of the assessment was not viable as a widely deployable diagnostic tool. However, this experiment did provide insights that led to effective modifications for the second assessment. The high prevalence of side biases among dogs that failed to complete the test suggested that perseverative tendencies may be strong in dogs. This led to the addition of correction trials after incorrect training and refresher trials, as well as the analysis of age-related changes in side bias in Experiment II. The finding that two dogs stopped participating after the first incorrect choice on a delay trial indicated that refresher trials may not successfully re-incentivize participation for some dogs. However, the fact that this occurred
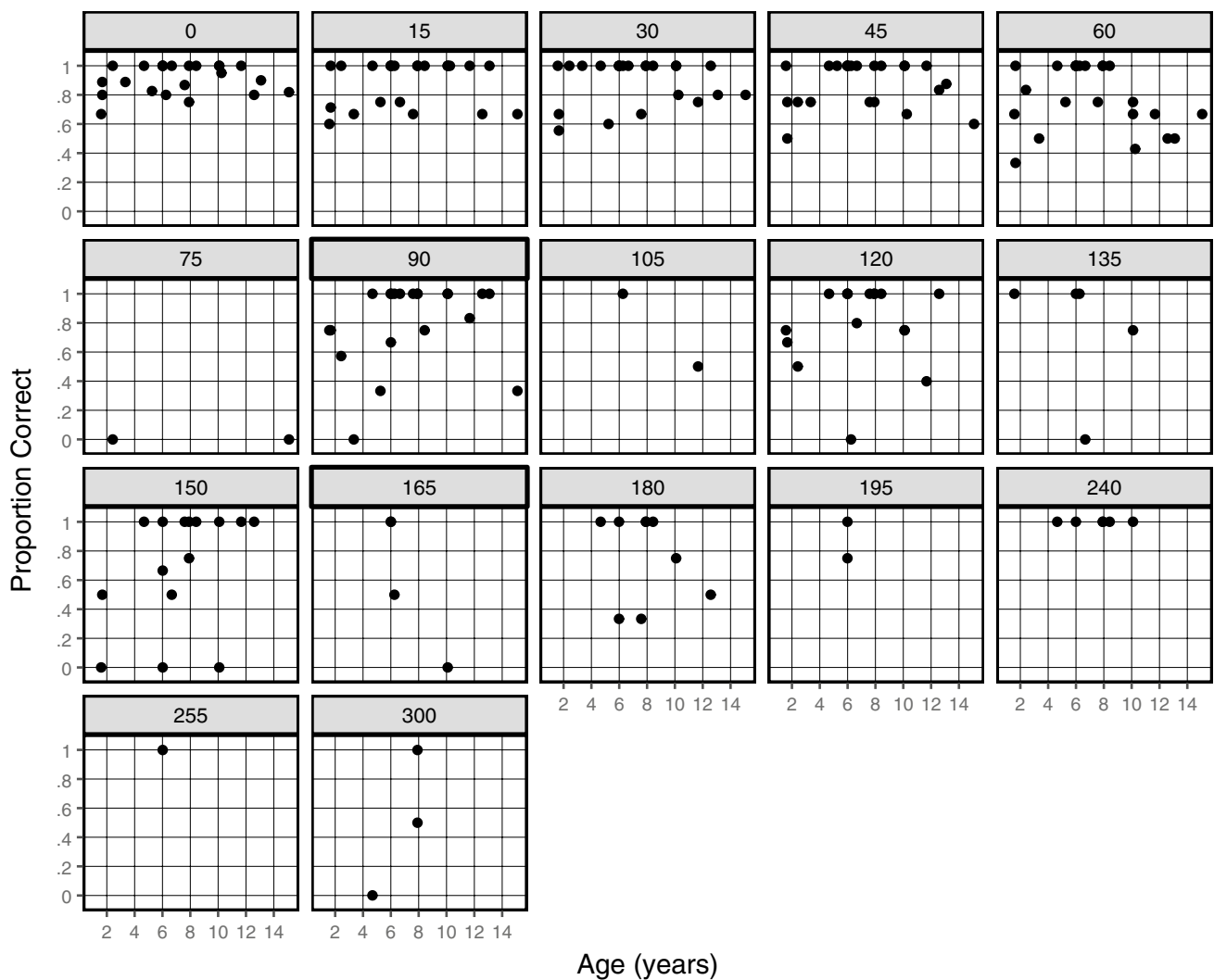
Fig. 9 Multi-plot of the proportion of correct choices (PCC) by age for each delay in Experiment II. Each panel shows PCC in the delay indicated in the gray header

only twice in 50 tests also suggested that, for most dogs, refresher trials may help to maintain participation throughout this hour-long assessment. Finally, the finding that two dogs did not participate in either experiment suggested that insufficient motivation may present a small but unavoidable obstacle to assessments for ARCD. However, the fourfold decrease in fail-out due to non-participation in Experiment II indicated that testing rules may be modified to greatly increase motivation and participation.

Experiment II—assessment validity

Proportion of correct choices

Dogs that completed the second assessment provided strong evidence of learning the spatial-matching rule. Indeed, only three dogs did not significantly differ from

chance in zero-delay trials and in each case, this was reflective of a small sample size (mean = 7) rather than poor performance (mean PCC = 81%). These dogs excelled in delay trials (PCC = 95%) and thus additional zero-delay trials (i.e., refreshers) were not needed to confirm acquisition of the spatial-matching rule. Moreover, the testing rules only allowed for such a small number of zero-delay trials when the dog committed few errors in either T3 or testing. Nonetheless, these non-significant results could be prevented by implementing additional refresher trials at the end of the test and this would require little additional testing time.

Evidence that dogs continued to use the spatial-matching rule to retrieve treats in delay trials was also robust. Only 4 of 23 dogs failed to choose the correct box more often than chance in delay trials but each succeeded in doing so in zero-delay trials. This indicates that incorrect choices in delay

trials resulted from insufficient SWM rather than failure to apply the spatial-matching rule.

Compared to previous SWM assessments for ARCD in colony beagles, this assessment required far fewer trials to confirm that the dog learned the spatial-matching task. Whereas beagles completed up to 40 daily sessions of ten training trials in previous assessments, dogs in the present study completed an average of 9.57 training trials and 10.77 refresher trials (see Fig. 6). This increase in efficiency was partially enabled by adopting a less cognitively demanding task. The delayed non-matching to position (DNMP) tasks used in previous studies required additional information processing and greater behavioral inhibition relative to the simple spatial-matching task used in the present study. Additionally, completing each training step in this staircase procedure required as few as two consecutive correct choices whereas DNMP training criteria required nine of ten, or eighteen of twenty correct choices. These lenient criteria and, in turn, rapid shaping and training procedures were made possible by the inclusion of response-dependent refresher trials, which collected additional evidence of the ability to follow the spatial-matching rule from any dog that did not provide this evidence in delay trials.

To our knowledge, refresher trials are the first controls in SWM assessments for dogs which assess task retention independently from task acquisition. A dog may stop following previously learned rules due to intertrial interference, the development of ineffective strategies, or fatigue. The dog may also meet training criteria by random chance and, therefore, begin the test before truly learning the reward-contingencies or appropriate responses. Traditional procedures do not address these possibilities and assume that the dog retained the task based solely on evidence that the dog learned the task in training. In the present study, dogs made correct choices in over 90% of the 283 refresher trials and the only dog below 80% PCC in refresher trials (3 of 4 correct) made correct choices in 20 of 21 delay trials. Thus, refresher trials provided evidence that each dog remembered the rules of the task and in turn assessed SWM.

Although training criteria in the present study were lenient relative to assessments for colony beagles, they provided substantial improvements relative to the two previous SWM assessments for ARCD in pet dogs. González-Martínez et al. (2013) compared the tendencies of old and young dogs to locate a hidden treat in an open field test after a 60-s retention period. They scored whether the dog (1) immediately located the treat, (2) located the treat after searching, (3) failed to locate the treat, or (4) made no attempt to locate the treat. Piotti et al. (2017) compared the tendencies of old and young dogs to locate a treat hidden in one of five containers after a 30-s retention period. They analyzed the first container approached by the dog in each of five trials (one per container).

Although both studies found significant differences between young and old dogs, neither confirmed that the test assessed SWM. The single-trial procedure used by González-Martínez et al. (2013) did not allow the dog to learn the task and the repeated-trials procedure used by Piotti et al. (2017) provided no incentives for learning the task given that the dog was allowed to retrieve the treat after an incorrect choice. In addition, neither study implemented training trials and neither assessed whether individual performance differed from random chance. Thus, these studies did not demonstrate that each dog encoded and remembered the location baited by the experimenter in the most recent demonstration. Furthermore, these studies did not confirm that all dogs were motivated to retrieve treats. In contrast, González-Martínez et al. (2013) scored non-participation as the most severe memory deficit. Ultimately, the present study may provide the first assessment for ARCD in pet dogs which directly measures SWM rather than incidental or uninformed search behavior (i.e., behavior which is likely influenced by SWM but to an unknown extent).

Delay-dependent performance

As illustrated by these previous assessments for pet dogs, errors in trials with delays can only be attributed to memory deficits if these errors are less frequent in trials without delays. Moreover, lower performance in delay trials than in training trials indicates that an assessment is memory dependent. Demonstrating that performance decreases with increasing retention periods provides even stronger evidence of a valid memory assessment. In the present study, the probability of choosing the correct box during the test decreased significantly with delay length. Although dogs only attempted longer delays in later trials, the effect of trial number was not significant and thus the effect of delay could not be attributed to fatigue or intertrial interference. In addition, differences in the delays and number of trials attempted by dogs of different ages could not account for this finding. Ultimately, this delay-dependent performance supports the validity of the present assessment as a test for SWM.

Our findings align with previous indications that spatial-matching paradigms may be used to assess SWM in pet dogs. Fiset et al. (2003) found that the spatial recall of young pet dogs in a similar three-location test gradually declined between 10 and 60 s, at which point performance remained stable and above chance for up to 240 s. Hunter (1913) and Walton (1915) observed delay-dependent performance in individual dogs using two- and three-location delayed-response tests. To our knowledge, all other previous SWM tests either used non-matching to position tasks or did not test for delay dependence.

Age-dependent performance

The finding that older dogs were less likely to choose correctly indicates that this memory assessment was sensitive to ARCD. Moreover, the outcomes from the GLMMs of test–trial outcome indicated that differences in the delays and number of trials attempted by dogs of different ages could not account for these deficits in older dogs. However, such models of trial-by-trial performance do not provide a means of determining whether dogs that complete this assessment in the future are cognitively impaired. To be clinically deployable ARCD assessments need to produce a single, easily interpreted score which pet owners and clinicians can compare either to previous tests on the same dog, or to a performance range typical of unimpaired dogs. Thus, the finding that cumulative memory scores declined after middle age may satisfy this important requirement of a widely deployable ARCD assessment.

This study also provided consistent evidence of memory deficits in young dogs. Each of the four dogs younger than 2.5 years was more likely to choose the incorrect box in a delay trial than each of the twelve dogs between 2.5 and 8.5 years of age. In addition, ten of the twelve middle-aged dogs achieved a higher CMS than the four young dogs. However, the age of peak performance (around 7 years) was likely overestimated given that the only dog between 2.5 and 5 years of age included in these analyses was by far the top performer in this study. This dog chose the correct box in 18 of 18 trials up to the 4-min delay, thereby achieving the only perfect CMS. Likewise, the other dog in this age range (excluded due to low participation) made correct choices in all of the delay trials in which it participated.

Given that the youngest dog tested in the present study was nearly an adult (age = 17 months), the curvilinear relationship between age and performance indicates that SWM develops slowly in pet dogs. Indeed, assessments that are sensitive to cognitive decline should also be sensitive to cognitive development. Although previous studies have primarily used decreasing linear functions to model ARCD in dogs, studies on non-human primates (Manrique and Call 2015), humans (Craik and Bialystok 2006; Harada et al. 2013), and dogs (Fox 1971) suggest that curvilinear models better estimate lifespan changes in cognitive abilities. In an analysis of citizen-science data from over 4000 pet dogs, Watowich et al. (2020) found that quadratic age functions better estimated performance in a broad range of cognitive assessments. Old and young dogs displayed deficits in social communication, inhibitory control, selective attention, reasoning, and working memory. Furthermore, these quadratic trends were strongest in tests for SWM and inhibitory control. In line with these findings, Gathercole et al. (2004) and Pickering et al. (2001) found that SWM continues to develop through adolescence and even early adulthood in humans.

Methodological differences may explain why previous experimental studies (i.e., studies which did not use citizen-science data) have not observed SWM deficits in young dogs. To our knowledge, eight such studies have examined the relationship between age and SWM in dogs. Two of these studies did not report the age ranges used to group dogs (Araujo et al. 2005; Piotti et al. 2017) and five of the other six studies used unique age groups. In addition, four studies tested laboratory-housed colony beagles (Chan et al. 2002; Tapp et al. 2003b; Araujo et al. 2005; Studzinski et al. 2006) and two tested both colony beagles and shelter-sourced dogs of uncertain ages (Head et al. 1995; Adams et al. 2000b). These six studies on beagles used five different variants of the DNMP task, all of which required additional information processing and greater behavioral inhibition relative to the simple matching task used in the present study. Two previous studies have tested pet dogs but neither confirmed that each individual was capable of and was motivated to complete the task (González-Martínez et al. 2013; Piotti et al. 2017). Ultimately, additional studies controlling for these differences are needed to clarify the developmental trajectory of SWM in dogs.

Although the present study is the first to observe SWM deficits in young dogs in a controlled experiment, previous studies have found little evidence that SWM declines monotonically with age from young adulthood onwards. Only three of these eight studies compared young and middle-aged dogs and their findings were inconsistent. Head et al. (1995) found that young and middle-aged beagles and pound-sourced dogs (of uncertain age) did not significantly differ in a three-location DNMP task but performance declined with age at some delays. However, this study did not include any dogs between 3 and 7 years of age, and did not test for a nonmonotonic, curvilinear relationship between age and performance. Using the same task, Studzinski et al. (2006) found that puppies (< 1year) and young beagles (1–3 years) performed better than middle-aged beagles (6–8 years). However, only one of the thirteen middle-aged beagles was sourced from a population that was found to significantly outperform the other populations sampled in this study. In contrast, all of the puppies and one-third of the young beagles came from the high performing population. González-Martínez et al. (2013) found that young and middle-aged pet dogs did not significantly differ in an open-field food-search task but it is unclear whether this assessment measured SWM.

Although four previous studies found that young beagles outperformed old beagles, only two of these findings disagree with the present study. Young beagles in Chan et al. (2002) and Tapp et al. (2003b) were between 3–5 and 3.4–6.6 years of age, respectively. In the present study, dogs within these age ranges also outperformed older dogs. This highlights how comparisons between two age groups with

truncated age ranges may erroneously indicate that performance declines linearly with age. When performance peaks in midlife, age groups can be defined such that young dogs appear to perform better, worse, or the same as old dogs. Furthermore, at least three age groups are required to detect a curvilinear relationship. To avoid this confusion, we suggest that future studies on ARCD should analyze age as a continuous variable.

Experiment II—exploratory analyses

Perseveration

The significant relationship between age and side bias indicates that this assessment may be sensitive to selective attention and inhibitory control. Selective attention refers to an individual's ability to ignore irrelevant information to effectively process task-relevant information (McDowd and Oseas-Kreger 1991). Inhibitory control requires an individual to resist the impulse to act on irrelevant, distracting information (Bray et al. 2014). Thus, if the dog formed an association between a particular box and treats, the magnitude of this box preference may reflect its ability to selectively attended to the most recent demonstration or to inhibit the behavior of choosing the preferred box. In turn, the U-shaped relationship between age and side bias may indicate that inhibitory control and selective attention in pet dogs develop slowly and decline in old age.

Most studies on inhibitory control in pet dogs have observed performance deficits and stronger side biases in older individuals (e.g., Mongillo et al. 2013; Piotti et al. 2018) but similar studies on colony beagles have produced inconsistent results (Tapp et al. 2003a; Chan et al. 2002; but see Head et al. 1998). However, only one study has, to our knowledge, explicitly examined the development of these cognitive processes in dogs. In line with the present study, Wallis et al. (2014) found that selective attention and sensorimotor coordination in Border collies peaked in midlife. In humans, tasks that require selective attention and inhibitory control are particularly difficult for adolescents (Diamond 2013). Development of the prefrontal cortex, which may play a critical role in inhibitory control (Braver et al. 2001; Diamond 1990; Ridderinkhof et al. 2004), is also relatively slow in humans (Bunge et al. 2002; Dempster 1992; West 1996). Tapp et al. (2004) found that frontal lobe volume (as a proportion of total brain volume) was larger in young colony beagles (6 m—3.9 years) than in puppies (3 m). In contrast, total brain volume (as a proportion of total intracranial volume) did not differ between young dogs and puppies. Thus, neurological development also suggests that inhibitory control may develop slowly in pet dogs.

The exploratory analysis of side bias cannot rule out a number of alternative explanations. For example, when the dog failed to remember which box was baited, it may have defaulted to choosing a particular box. If so, incorrect choices would only occur when the dog failed to remember which box was baited and the preferred box happened to be incorrect in that trial. Thus, the magnitude of the dog's box preference may measure memory rather than inhibitory control, and the outcome of the side-bias analysis may provide additional evidence that middle-aged dogs were more likely to remember the location of the correct box. Alternatively, this finding may indicate that middle-aged dogs were less likely to choose a default box when they failed to remember the baited location. Ultimately, this memory assessment was not designed to measure side biases or the processes through which they develop. Therefore, future assessments using this two-location paradigm which manipulate the side baited by the experimenter in zero-delay trials may help to clarify the relationships between age, side biases, selective attention, and inhibitory control in pet dogs.

Head and body orientations

The findings of the present study indicate that dogs did not orient towards baited locations to circumvent the memory requirements of this spatial-matching task. Specifically, neither head nor body orientation towards the correct box during the delay improved performance. In contrast, dogs oriented their head and body away from the correct box in over 95% and 75% of all delay trials, respectively, but still chose the correct box in nearly 80% of these trials. In line with these findings, Fiset et al. (2000) found that the probability of a correct choice did not depend on the dog's head or body orientation immediately, 5 s, or 10 s after a reward was hidden in one of three boxes. Given that we did not code eye-tracking, it is possible that dogs visually fixated on the correct box. However, in a four-location visual displacement task, Gagnon and Doré (1994) found that dogs interrupted their visual fixation on the target box in 96.7% of 10 s and 20 s delay trials. Thus, visual fixation in delays of up to 300 s in the present study was unlikely.

The finding that dogs were more likely to choose the correct box when they remained oriented towards the testing room during the delay indicates that the ability to remain engaged may constrain the processes through which dogs maintain locations in recent memory. Humans attend to relevant spatial locations to maintain visuospatial information in working memory (Smyth and Scholey 1994) and thus, selective spatial attention serves as a rehearsal mechanism for SWM (Awh et al. 1998). In consequence, human observers respond faster and more accurately to stimuli within the spatial region of their current attention (Posner 1980). In the present study, the orientation of the dog's body was used to measure its attentional space. Therefore, dogs may have ceased to actively maintain the treat location in memory

when they lost interest in the test or were distracted (i.e., when they turned away from the testing room).

Such disruptions likely occur more frequently over increasing lengths of time. Thus, this increasing difficulty of remaining engaged may explain the finding that performance decreased with increasing delay. Indeed, after accounting for forward body orientation, the effect of delay on delay-trial outcome was only marginally significant. In contrast, the effect of delay on test–trial outcome was highly significant. However, this weaker prediction may be attributed to the fact that the latter model did not account for the higher performance of dogs in zero-delay trials.

These competing interpretations highlight the exploratory nature of the analyses of orientation behaviors. Moreover, an interruption to forward body orientation may be a consequence rather than a cause of a failure to remember the correct location. In addition, forward head orientation during the delay, the other hypothesized indication of sustained engagement, did not improve performance. Ultimately, additional studies are needed to clarify the relationship between head and body orientations, engagement, retention periods, and SWM.

Delay optimization

Future versions of this assessments may be able to detect ARCD more rapidly using only 45 s and 60 s delays. Delays which are so short that they present little challenge to cognitively impaired dogs may only help to detect the most severe ARCD. Given that only two of eight old dogs (≤ 8 years) displayed deficits at the 15 s delay, and that age did not predict performance at the 15 s or 30 s delays, these delays may not be worth the additional testing time in future assessments. Given that AGE and AGE^2 accounted for the largest proportion of variation in performance at the 45 s and 60 s delays, the inclusion of longer delays in future assessments may also prolong the test unnecessarily while reducing sensitivity to ARCD.

Conclusions

In this study, we designed an adaptive staircase assessment for SWM in pet dogs that were sensitive to ARCD. Both young and old dogs displayed performance deficits relative to middle-aged dogs indicating that SWM not only declines in old age but also may develop slowly in young, adult dogs. Orientations of the dog's head and body suggested that the ability to remain engaged with the task may constrain the dog's ability to maintain locations in recent memory. However, dogs did not orient towards baited locations to circumvent the memory requirements of this task. Up-down psychometric staircase methods and,

more broadly, adaptive testing procedures may be used to efficiently assess the cognitive abilities of pet dogs within a single session. However, future assessments which adopt this paradigm may be able to detect ARCD more rapidly by testing SWM in only 45-s and 60-s delay trials. Alternatively, this paradigm may be used to detect age-related deficits in selective attention and inhibitory control.

Compliance with ethical standards

References

Adams B, Chan A, Callahan H, Milgram NW (2000) The canine as a model of human cognitive aging: recent developments. Prog Neuropsychopharmacol Biol Psychiatry 24(5):675–692. https://doi.org/10.1016/S0278-5846(00)00101-9

Adams B, Chan A, Callahan H, Siwak C, Tapp D, Ikeda-Douglas C, Atkinson P, Head E, Cotman CW, Milgram NW (2000) Use of a delayed non-matching to position task to model age-dependent cognitive decline in the dog. Behav Brain Res 108(1):47–56. https://doi.org/10.1016/S0166-4328(99)00132-1

Araujo JA, Studzinski CM, Milgram NW (2005) Further evidence for the cholinergic hypothesis of aging and dementia from the canine model of aging. Prog Neuropsychopharmacol Biol Psychiatry 29(3):411–422. https://doi.org/10.1016/j.pnpbp.2004.12.008

Araujo JA, Baulk J, de Rivera C (2017) The aged dog as a natural model of alzheimer's disease progression. In: Landsberg G, Maďari A, Žilka N (eds) Canine and Feline Dementia. Springer, Cham, pp 69–94

Awh E, Jonides J, Reuter-Lorenz PA (1998) Rehearsal in spatial working memory. J Exp Psychol Hum Percept Perform 24(3):780

Azkona G, García-Belenguer S, Chacón G, Rosado B, León M, Palacio J (2009) Prevalence and risk factors of behavioural changes associated with age-related cognitive impairment in geriatric dogs. J Small Anim Pract 50(2):87–91. https://doi.org/10.1111/j.1748-5827.2008.00718.x

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

Brambor T, Clark WR, Golder M (2006) Understanding interaction models: improving empirical analyses. Political analysis. Political analysis, 63–82

Braver TS, Barch DM, Keys BA, Carter CS, Cohen JD, Kaye JA, Janowsky JS, Taylor SF, Yesavage JA, Mumenthaler MS, Jagust WJ (2001) Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. J Exp Psychol Gen 130(4):746. https://doi.org/10.1037/0096-3445.130.4.746

Bray EE, MacLean EL, Hare BA (2014) Context specificity of inhibitory control in dogs. Anim Cogn 17(1):15–31. https://doi.org/10.1007/s10071-013-0633-z

Bunge SA, Dudukovic NM, Thomason ME, Vaidya CJ, Gabrieli JD (2002) Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. Neuron 33(2):301–311. https://doi.org/10.1016/S0896-6273(01)00583-9

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. https://doi.org/10.1037/0735-7044.116.3.443

Chapagain D, Range F, Huber L, Virányi Z (2018) Cognitive aging in dogs. Gerontol 64(2):165–171. https://doi.org/10.1159/000481621

Cornsweet TN (1962) The staircrase-method in psychophysics. Am J Psychol 75(3):485–491. https://doi.org/10.2307/1419876

Craik FI, Bialystok E (2006) Cognition through the lifespan: mechanisms of change. Trends Cogn Sci 10(3):131–138. https://doi.org/10.1016/j.tics.2006.01.007

Dempster FN (1992) The rise and fall of the inhibitory mechanism: toward a unified theory of cognitive development and aging. Dev Rev 12(1):45–75. https://doi.org/10.1016/0273-2297(92)90003-K

Diamond A (1990) Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. Ann N Y Acad Sci 608(1):637–676. https://doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond A (2013) Executive functions. Annu Rev Psychol 64:135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Ehrenstein WH, Ehrenstein A (1999) Psychophysical methods. In: Windhorst U, Johansson H (eds) Modern techniques in neuroscience research. Springer, Berlin, pp 1211–1241

Fast R, Schütt T, Toft N, Møller A, Berendt M (2013) An observational study with long-term follow-up of canine cognitive dysfunction: clinical characteristics, survival, and risk factors. Vet Intern Med 27(4):822–829. https://doi.org/10.1111/jvim.12109

Fiset S, Gagnon S, Beaulieu C (2000) Spatial encoding of hidden objects in dogs (*Canis familiaris*). J Comp Psychol 114(4):315. https://doi.org/10.1037/0735-7036.114.4.315

Fiset S, Beaulieu C, Landry F (2003) Duration of dogs' (*Canis familiaris*) working memory in search for disappearing objects. Anim Cogn 6(1):1–10. https://doi.org/10.1007/s10071-002-0157-4

Fox MW (1971) Integrative development of brain and behavior in the dog. University of Chicago Press, Chicago

Gagnon S, Doré FY (1994) Cross-sectional study of object permanence in domestic puppies (*Canis familiaris*). J Comp Psychol 108(3):220. https://doi.org/10.1037/0735-7036.108.3.220

Gathercole SE, Pickering SJ, Ambridge B, Wearing H (2004) The structure of working memory from 4 to 15 years of age. Dev Psychol 40(2):177. https://doi.org/10.1037/0012-1649.40.2.177

Gilmore KM, Greer KA (2015) Why is the dog an ideal model for aging research? Exp Gerontol 71:14–20. https://doi.org/10.1016/j.exger.2015.08.008

González-Martínez Á, Rosado B, Pesini P, García-Belenguer S, Palacio J, Villegas A, Suárez ML, Santamarina G, Sarasa M (2013) Effect of age and severity of cognitive dysfunction on two simple tasks in pet dogs. Vet J 198(1):176–181. https://doi.org/10.1016/j.tvjl.2013.07.004

Greer KA, Canterberry SC, Murphy KE (2007) Statistical analysis regarding the effects of height and weight on life span of the domestic dog. Res Vet Sci 82(2):208–214. https://doi.org/10.1016/j.rvsc.2006.06.005

Harada CN, Love MCN, Triebel K (2013) Normal cognitive aging. Clin Geriatr Med 29:737–752. https://doi.org/10.1016/j.cger.2013.07.002

Head E (2001) Brain aging in dogs: parallels with human brain aging and Alzheimer's disease. Vet Ther 2(3):247–260

Head E (2013) A canine model of human aging and Alzheimer's disease. Biochim Biophys Acta Mol Basis Dis 1832(9):1384–1389. https://doi.org/10.1016/j.bbadis.2013.03.016

Head E, Mehta R, Hartley J, Kameka M, Cummings BJ, Cotman CW, Ruehl WW, Milgram NW (1995) Spatial learning and memory as a function of age in the dog. Behav Neurosci 109(5):851

Head E, Callahan H, Muggenburg BA, Cotman CW, Milgram NW (1998) Visual-discrimination learning ability and β-amyloid accumulation in the dog. Neurobiol Aging 19(5):415–425. https://doi.org/10.1016/S0197-4580(98)00084-0

Hunter WS (1913) The delayed reaction in animals and children. Behav Monographs 2:1–85. https://doi.org/10.1037/0735-7044.109.5.851

Kaeberlein M, Creevy KE, Promislow DE (2016) The dog aging project: translational geroscience in companion animals. Mamm Genome 27(7–8):279–288. https://doi.org/10.1007/s00335-016-9638-7

Kaernbach C (1991) Simple adaptive testing with the weighted up-down method. Percept Psychophys 49(3):227–229

Kensinger EA, Shearer DK, Locascio JJ, Growdon JH, Corkin S (2003) Working memory in mild Alzheimer's disease and early Parkinson's disease. Neuropsychol 17(2):230. https://doi.org/10.1037/0894-4105.17.2.230

Kraus C, Pavard S, Promislow DE (2013) The size-life span trade-off decomposed: why large dogs die young. Am Nat 181:492–505. https://doi.org/10.1086/669665

Landsberg GM, Nichol J, Araujo JA (2012) Cognitive dysfunction syndrome: a disease of canine and feline brain aging. Vet Clin North Am Small Anim Pract 42(4):749–768. https://doi.org/10.1016/j.cvsm.2012.04.003

Madari A, Farbakova J, Katina S, Smolek T, Novak P, Weissova T, Novak M, Zilka N (2015) Assessment of severity and progression of canine cognitive dysfunction syndrome using the CAnine DEmentia Scale (CADES). Appl Anim Behav Sci 171:138–145. https://doi.org/10.1016/j.applanim.2015.08.034

Manrique HM, Call J (2015) Age-dependent cognitive inflexibility in great apes. Anim Behav 102:1–6. https://doi.org/10.1016/j.anbehav.2015.01.002

Marinissen EJ, Singh A, Glotter D, Esposito M, Carulli JM, Nahar A, Butler KM, Appello D, Portelli C (2010) Adapting to adaptive testing. In: 2010 design, automation and test in Europe conference and exhibition. IEEE, Dresden 556–561

McDowd JM, Oseas-Kreger DM (1991) Aging, inhibitory processes, and negative priming. J Gerontol 46(6):P340–P345. https://doi.org/10.1093/geronj/46.6.P340

Milgram NW, Adams B, Callahan H, Head E, Mackay B, Thirlwell C, Cotman CW (1999) Landmark discrimination learning in the dog. Learn Mem 6(1):54–61. https://doi.org/10.1101/lm.6.1.54

Mongillo P, Araujo JA, Pitteri E, Carnier P, Adamelli S, Regolin L, Marinelli L (2013) Spatial reversal learning is impaired by age in pet dogs. Age 35(6):2273–2282. https://doi.org/10.1007/s11357-013-9524-0

Mongillo P, Scandurra A, Aniello BD, Marinelli L (2017) Effect of sex and gonadectomy on dogs' spatial performance. Appl Anim Behav Sci 191:84–89. https://doi.org/10.1016/j.applanim.2017.01.017

Pickering SJ (2001) The development of visuo-spatial working memory. Memory 9(4–6):423–432. https://doi.org/10.1080/09658210143000182

Piotti P, Szabó D, Wallis L, Bognár Z, Stiegmann B, Egerer A, Marty P, Kubinyi E (2017) The effect of age on visuo-spatial short-term memory in family dogs. Pet Behav Sci 4:17–19. https://doi.org/10.21071/pbs.v0i4.10130

Piotti P, Szabó D, Bognár Z, Egerer A, Hulsbosch P, Carson RS, Kubinyi E (2018) Effect of age on discrimination learning, reversal learning, and cognitive bias in family dogs. Learn Behav 46(4):537–553. https://doi.org/10.3758/s13420-018-0357-7

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32(1):3–25

Ridderinkhof KR, Van Den Wildenberg WP, Segalowitz SJ, Carter CS (2004) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn 56(2):129–140. https://doi.org/10.1016/j.bandc.2004.09.016

Rofina JE, Van Ederen AM, Toussaint MJ, Secreve M, Van Der Spek A, Van Der Meer I, Van Eerdenburg FJ, Gruys E (2006) Cognitive disturbances in old dogs suffering from the canine counterpart of Alzheimer's disease. Brain Res 1069(1):216–226. https://doi.org/10.1016/j.brainres.2005.11.021

Salvin HE, McGreevy PD, Sachdev PS, Valenzuela MJ (2010) Under diagnosis of canine cognitive dysfunction: a cross-sectional survey of older companion dogs. Vet J 184(3):277–281. https://doi.org/10.1016/j.tvjl.2009.11.007

Salvin HE, McGreevy PD, Sachdev PS, Valenzuela MJ (2012) The effect of breed on age-related changes in behavior and disease prevalence in cognitively normal older community dogs, *Canis lupus familiaris*. J Vet Behav 7:61–69. https://doi.org/10.1016/j.jveb.2011.06.002

Simone PM, Baylis GC (1997) Selective attention in a reaching task: effect of normal aging and Alzheimer's disease. J Exp Psychol Hum Percept Perform 23(3):595. https://doi.org/10.1037/0096-1523.23.3.595

Siwak-Tapp CT, Head E, Muggenburg BA, Milgram NW, Cotman CW (2007) Neurogenesis decreases with age in the canine hippocampus and correlates with cognitive function. Neurobiol Learn Mem 88(2):249–259. https://doi.org/10.1016/j.nlm.2007.05.001

Smyth MM, Scholey KA (1994) Interference in immediate spatial memory. Mem Cognit 22(1):1–3. https://doi.org/10.3758/BF03202756

Solt F, Hu Y, Kenke B (2015) Interplot: Plot the effects of variables in interaction terms. Available at the Comprehensive R Archive Network (CRAN)

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 Mem 86(2):197–204. https://doi.org/10.1016/j.nlm.2006.02.005

Szabó D, Gee NR, Miklósi Á (2016) Natural or pathologic? Discrepancies in the study of behavioral and cognitive signs in aging family dogs. J Vet Behav 11:86–98. https://doi.org/10.1016/j.jveb.2015.08.003

Szabó D, Miklósi Á, Kubinyi E (2018) Owner reported sensory impairments affect behavioural signs associated with cognitive decline in dogs. Behav Process 157:354–360. https://doi.org/10.1016/j.beproc.2018.07.013

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. Lear Mem 10(1):64–73. https://doi.org/10.1101/lm.54403

Tapp PD, Siwak CT, Estrada J, Holowachuk D, Milgram NW (2003) Effects of age on measures of complex working memory span in the beagle dog (*Canis familiaris*) using two versions of a spatial list learning paradigm. Learn Mem 10(2):148–160. https://doi.org/10.1101/lm.56503

Tapp PD, Siwak CT, Gao FQ, Chiou JY, Black SE, Head E, Muggenburg BA, Cotman CW, Milgram NW, Su MY (2004) Frontal lobe volume, function, and β-amyloid pathology in a canine model of aging. J Neurosci 24(38):8205–8213. https://doi.org/10.1523/JNEUROSCI.1339-04.2004

Vite CH, Head E (2014) Aging in the canine and feline brain. Vet Clin Small Anim Pract 44(6):1113–1129. https://doi.org/10.1016/j.cvsm.2014.07.008

Wainer H, Lewis C (1990) Toward a psychometrics for testlets. J Educ Meas 27(1):1–14. https://doi.org/10.1111/j.1745-3984.1990.tb00730.x

Wallis LJ, Range F, Müller CA, Serisier S, Huber L, Virányi Z (2014) Lifespan development of attentiveness in domestic dogs: drawing parallels with humans. Front Psychol 5:71. https://doi.org/10.3389/fpsyg.2014.00071

Wallis LJ, Virányi Z, Müller CA, Serisier S, Huber L, Range F (2016) Aging effects on discrimination learning, logical reasoning and memory in pet dogs. Age 38(1):6. https://doi.org/10.1007/s11357-015-9866-x

Walton AC (1915) The influence of diverting stimuli during delayed reaction in dogs. J Anim Behav 5:259–291. https://doi.org/10.1037/h0074449

Watowich MM, MacLean EL, Hare B, Call J, Kaminski J, Miklósi Á, Snyder-Mackler N (2020) Age influences domestic dog cognitive performance independent of average breed lifespan. Anim Cogn 23:795–805. https://doi.org/10.1007/s10071-020-01385-0

Watson AB, Fitzhugh A (1990) The method of constant stimuli is inefficient. Percept Psychophys 47(1):87–91. https://doi.org/10.3758/BF03208169

West RL (1996) An application of prefrontal cortex function theory to cognitive aging. Psychol Bull 120(2):272. https://doi.org/10.1111/jvim.12109