GeroScience
https://doi.org/10.1007/s11357-024-01278-x

ORIGINAL ARTICLE

Characterizing dog cognitive aging using spontaneous problem-solving measures: development of a battery of tests from the Dog Aging Project

Stephanie H. Hargrave · Emily E. Bray · Stephanie McGrath · Gene E. Alexander · Theadora A. Block · Naomi Chao · Martin Darvas · Laura E. L. C. Douglas · Janet Galante · Brenda S. Kennedy · Breonna Kusick · Julie A. Moreno · Daniel E. L. Promislow · David A. Raichlen · Lorelei R. Switzer · Lily Tees · Mikayla Underwood Aguilar · Silvan R. Urfer · Dog Aging Project Consortium · Evan L. MacLean

Received: 24 January 2024 / Accepted: 1 July 2024

Abstract Companion dogs are a valuable model for aging research, including studies of cognitive decline and dementia. With advanced age, some dogs spontaneously develop cognitive impairments and neuropathology resembling features of Alzheimer's disease. These processes have been studied extensively in laboratory beagles, but the cognitive assays used in that context—which rely on time-consuming operant procedures—are not easily scalable to large samples of community-dwelling companion dogs. We developed a battery of five short-form tasks targeting three aspects of cognition that are impaired in Alzheimer's disease: spatial memory, executive functions, and social cognition. In Experiment 1, we tested a cross-sectional sample of dogs ($N = 123$) and estimated associations between age and task performance. Older dogs scored lower on measures of spatial learning, memory, and response flexibility, and spent less time near, but more time gazing at, the experimenter. We found no differences in associations between age and performance across dogs of different body masses, a proxy for expected lifespan. In Experiment 2, we demonstrated the feasibility

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11357-024-01278-x.

S. H. Hargrave · E. E. Bray · G. E. Alexander · E. L. MacLean (✉)
Department of Psychology, University of Arizona, Tucson, AZ, USA
e-mail: evanmaclean@arizona.edu

S. H. Hargrave · E. E. Bray · N. Chao · L. R. Switzer · E. L. MacLean
College of Veterinary Medicine, University of Arizona, Tucson, AZ, USA

E. E. Bray · T. A. Block · L. E. L. C. Douglas · B. S. Kennedy
Canine Companions, National Headquarters, Santa Rosa, CA, USA

S. McGrath · B. Kusick
Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

G. E. Alexander
BIO5 Institute, University of Arizona, Tucson, AZ, USA

G. E. Alexander
Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA

G. E. Alexander
Department of Psychiatry, University of Arizona, Tucson, AZ, USA

G. E. Alexander
Neuroscience Graduate Interdisciplinary Program, University of Arizona, Tucson, AZ, USA

G. E. Alexander
Physiological Sciences Graduate Interdisciplinary Program, University of Arizona, Tucson, AZ, USA

Published online: 06 August 2024

of these measures in clinical settings ($N=35$). Dogs meeting clinical criteria for moderate or severe cognitive impairment scored lower, on average, than dogs characterized as mildly impaired and healthy agers, although these distributions overlapped. However, few dogs in our study cohort met the criteria for moderate or severe impairment. The measures presented here show promise for deployment in large-scale longitudinal studies of companion dogs, such as the Dog Aging Project.

Keywords Alzheimer's disease · Cognitive aging · Dogs

Introduction

Companion dogs are a valuable model for translational aging research due to a wealth of features shared with humans, including high levels of genetic and phenotypic diversity, exposure to human living environments and their various associated risk factors, highly variable lifestyles, comparable disease risks and burdens, and access to a sophisticated healthcare system with associated medical records [1, 2]. However, in contrast to humans, dogs have much shorter lifespans, facilitating efficient longitudinal studies of the genetic, nutritional, and environmental determinants of aging [3].

Dogs have also been proposed as a valuable natural model for Alzheimer's disease. Nonhuman animal models of Alzheimer's disease can play an important role in understanding the pathogenesis of dementia and are critical for preclinical studies of novel therapeutics. However, the translational potential of animal models depends on the extent to which they replicate core features of Alzheimer's disease and its etiology. Currently, no animal model fully captures the clinical and neurodegenerative components of Alzheimer's disease, and reliance on genetically homogenous populations studied in lab environments limits translational potential [4]. With advanced age, a subset of dogs spontaneously develops aspects of neuropathology (e.g., amyloid-beta plaques) and impairments in learning, memory, and executive functions that closely resemble those observed in Alzheimer's disease [2, 5–12]. Diagnostic tools for this condition—canine cognitive dysfunction—remain limited, and it is believed that canine cognitive dysfunction is profoundly underdiagnosed [13]. Current veterinary diagnostic tools rely heavily on owner reports and may only be sensitive to advanced stages of dementia, manifested through behavioral changes such as aimless wandering, disorientation, loss of housetraining, altered sleep/wake cycles, and inability to recognize familiar individuals [9, 14, 15].

Laboratory studies suggest that, as in Alzheimer's disease, these late-stage symptoms of dementia are preceded by mild cognitive impairments, which begin in middle age and are detectable

G. E. Alexander
Arizona Alzheimer's Consortium, Phoenix, AZ, USA

M. Darvas · D. E. L. Promislow · S. R. Urfer
Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA

J. Galante
Sit! Stay! Play!, Tucson, AZ, USA

J. A. Moreno
Department of Environmental and Radiological Health Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

D. E. L. Promislow
Department of Biology, University of Washington, Seattle, WA, USA

D. E. L. Promislow
Current affiliation: Jean Mayer USDA Human Nutrition Research Center On Aging, Tufts University, Boston, MA, USA

D. A. Raichlen
Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA

D. A. Raichlen
Department of Anthropology, University of Southern California, Los Angeles, CA, USA

L. Tees
College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

M. Underwood Aguilar
School of Veterinary Medicine, University of California-Davis, Davis, CA, USA

through neuropsychological testing [6]. Detection of these initial changes is critical given that early intervention is widely believed to present the best opportunity for delaying or preventing the progression of dementia [16–18]. Early detection is also paramount for the design of studies testing novel interventions or therapeutics; however, existing tools are limited in their ability to capture mild cognitive impairments in dogs. Therefore, in order for us to leverage companion dogs as a model of Alzheimer's disease and cognitive aging more broadly, objective, easily administered cognitive assays, which are sensitive to the initial stages of cognitive decline, are essential. Although a range of functional assays have been developed and validated in studies with laboratory beagles, these tests rely heavily on operant conditioning, requiring lengthy training and testing protocols that are not feasible with large samples of privately owned companion dogs [3]. In recent decades, comparative psychologists have invested heavily in the development of cognitive tasks that rely on spontaneous problem-solving rather than operant procedures, enabling rapid data acquisition in large samples [19–23]. This approach has shown promise for studies of cognitive aging in companion dogs [24–27], but there are currently few well-validated problem-solving assays available to the research community.

Piotti et al. [28] reported that dogs learned location-based discriminations faster and more readily than those based on other stimulus characteristics, suggesting that location-based tasks may be particularly good candidates for short-format assessments of canine cognition. Several spatial reversal tasks have shown promise in measuring cognitive differences associated with age; older dogs consistently make more errors on such tasks [29, 30]. While such results are indicative of impaired response flexibility in older dogs, existing reversal learning assays still pose challenges for integration into short-format, multiple-task batteries due to relatively long task-durations, and/or retention measures collected weeks following initial learning.

Other recent studies have developed tasks to investigate the effects of age on memory and other aspects of cognition and behavior in dogs, such as problem-solving ability, neophilia, and social interaction. Chapagain et al. [31] found that older dogs performed worse on problem-solving tasks, exhibited less manipulative persistency with a food dispensing toy, were less successful on a detour task, were less playful, directed less attention towards and exhibited less attachment to their owners, were less willing to interact with a human stranger, and took longer to find hidden food during a memory task. Similarly, Piotti et al. [26] found that older dogs performed worse on a memory task requiring dogs to locate a hidden food reward after a 30-s distraction period.

Informed by this prior research, we developed a new, efficient, and easily administered neuropsychological assessment for age-related cognitive impairments in companion dogs. The tasks included in this battery focus on cognitive and behavioral processes implicated in both canine cognitive dysfunction and Alzheimer's disease: spatial memory, executive function, and social behavior [5, 6, 32–39]. Dogs were tested with five spontaneous problem-solving measures. Our primary aim was to assess associations between dog age and task performance in a diverse sample of companion dogs. Additionally, we conducted exploratory analyses to assess whether associations between age and cognitive performance vary by dog body mass (as a proxy for expected lifespan). Lastly, in a second experiment, we conducted a preliminary study to assess the feasibility of using these measures in clinical studies, including with dogs characterized as having cognitive impairments based on the CAnine DEmentia Scale (CADES) [14].

Experiment 1

Methods

Participants and data collection sites

Data collection occurred at three sites including a university research laboratory (Arizona Canine Cognition Center, Tucson, AZ, USA), a dog daycare facility (Sit! Stay! Play!, Tucson, AZ, USA), and a service dog training center (Canine Companions, Santa Rosa, CA, USA). Demographic information for participating dogs is provided in Table 1. Dogs tested at the Arizona Canine Cognition Center were recruited from a database of local dog owners who had expressed interest in participation in cognitive and behavioral studies. Dogs tested at the dog daycare facility were recruited through informational flyers provided to

Table 1 Selected characteristics of companion dog participants in Experiment 1. *IQR* interquartile range

	$N = 123^1$
Sex	
Intact female	9 (7.32%)
Intact male	3 (2.44%)
Neutered male	57 (46.34%)
Spayed female	54 (43.90%)
Breed category	
Mixed breed	66 (53.66%)
Purebred	57 (46.34%)
Test location	
Arizona Canine Cognition Center	46 (37.40%)
Canine Companions	30 (24.39%)
Sit! Stay! Play!	47 (38.21%)
Weight (kg)	24.94 (14.29, 29.02)
Age (years)	6.31 (2.96, 11.17)

[1] *n* (%); median (IQR)

clients. Dogs tested at the service dog training center were recruited through word of mouth and electronic invitations to dog owners near the data collection site.

General methods

Dogs participated in a battery of five tasks, typically administered across two sessions on separate days, with each session lasting approximately 30–60 min. The first session consisted of two memory tasks (two-location task, delayed search task) and a sensory screen, and the second session consisted of two measures of executive function (cylinder task, spatial reversal task) as well as a measure of social interaction with an unfamiliar experimenter (human interaction). Detailed methods for all tasks, presented in their order of administration, are described in dedicated subsections below. All procedures were approved by the University of Arizona Institutional Animal Care and Use Committee (protocol #16–175).

All tasks involved a handler who positioned the dog at a start location and timed the trials, and an experimenter who administered the test and recorded data. All sessions were video recorded. Dog responses were coded in real time for all tasks except the human interaction task, which was coded from video. Additionally, a random subset of ≥ 20% of trials from all tasks were later coded from video by a second observer to assess inter-rater reliability (see "Statistical analysis").

All tasks except the human interaction task involved dogs engaging in spontaneous problem-solving for food rewards. Dogs were not systematically fasted prior to participation. However, owners were informed that it may be helpful to avoid feeding them within a few hours prior to testing and/or to give them a partial rather than a full meal ahead of testing, particularly if their dog was likely to be satiated by 40 small treats. Food rewards used throughout the battery were typically small pieces of commercial meat-based dog treats (Jerky Treats, distributed by Big Heart Pet Brand, Orrville, OH), but occasionally varied according to dog preferences or dietary restrictions. To confirm that dogs were unable to choose accurately based on olfactory cues, we implemented an odor control task with 20 dogs. Mean performance in this task closely approximated chance expectation (null expectation = 25%, performance = 24%, 90% CI = 0.18%, 0.30%; see supplemental materials).

A total of ten experimenters administered behavioral tests in Experiment 1. Some experimenters were trained prior to the beginning of the study, while others were trained on an ongoing basis as needed to support ongoing data collection. New experimenters underwent a detailed training process which included implementing the tasks with a non-study dog while receiving feedback from a current experimenter and being supervised by existing experimenters the first time they performed the experimenter role with a study dog.

Across all tasks, if a dog did not move from their starting position when the experimenter said "okay" (the cue to release the dog), the experimenter repeated the "okay" release cue at 3-s intervals, up to a total of three times. If the dog had not initiated a response by third release cue (e.g., moving forward to search), the handler gently nudged the dog from behind to encourage them to move forward from the start line without directing them towards any particular choice location (memory tasks) or side (cylinder and spatial reversal). If at any point a dog became fearful or anxious during testing and was not easily soothed, we discontinued data collection for that task. If a dog exhibited fear, anxiety, or stress at the beginning of a session and did not become comfortable within approximately 10–15 min of acclimation to the room and experimenters, data collection was discontinued with this

subject ($N=3$). Likewise, if a dog was hesitant to consume the food rewards offered freely during acclimation, they also did not complete tasks ($N=3$).

At the Arizona Canine Cognition Center, owners were present (positioned in a neutral location behind the dog during testing) in addition to the experimenter and handler, but owners were typically absent during testing at the other locations. A minority of dogs who were tested at the Arizona Canine Cognition Center were uncomfortable being handled by an unfamiliar person but were otherwise motivated and capable of participating. In these cases ($N=10$), and for one dog tested at Canine Companions, the dog's owner performed the handler role (aligning the dog at the start position and releasing the dog when cued by the experimenter).

Lastly, some dogs tested at the daycare required brief acclimation sessions to become sufficiently comfortable with the experimenter and testing space ($N=11$). During these acclimation sessions, dogs were provided with food rewards and given opportunities to interact with research personnel and to explore the testing room but were not exposed to any of the test procedures.

Sensory screen Since older dogs may suffer from age-related hearing or vision loss, we conducted a brief sensory screening to assess responsiveness to visual and auditory stimuli. Failure on the vision screening was treated as an exclusion criterion given that all tasks required dogs to process visual stimuli. Failure on the auditory screening was not used as an exclusion criterion given that processing of auditory stimuli was not essential for task participation. In the vision screening, the experimenter stood 1 m in front of the dog and dropped a cotton ball (25 cm diameter) from the dog's eye level to the floor in front of the dog. The experimenter then kneeled 1 m in front of the dog and 0.67 m to the side, and using a finger, flicked the cotton ball such that it passed in front of and to the opposite side of the dog. The dog passed the visual screening if they followed the motion of the cotton ball with their head and/or eyes. All dogs passed the vision screening. The auditory screening consisted of an experimenter pressing a training clicker on the opposite side of their body from the dog such that the clicker was obscured from the dog's view and approximately 0.33 m from the dog's head. The dog passed the auditory screening if they showed an immediate (within 1 s) observable response to the clicker on either side, including but not limited to a rapid change in head orientation or an ear flick on the side of the stimulus. Six dogs did not pass the auditory screening.

The procedures described below are presented in order of administration.

Session 1

Blinds familiarization The two memory tasks involved dogs searching for food rewards behind v-shaped blinds (Figs. S1 and S2). Prior to implementing these tasks, we conducted a familiarization procedure to acclimate dogs to the testing environment, blinds, and process of searching for food behind the blinds. A single blind was positioned in a central location and the experimenter visibly placed a treat behind the blind as the dog watched. The dog was then immediately released to search for the food and was led to the treat by the experimenter if they did not quickly find it on their own. This process was repeated until the dog reliably searched for and located a treat placed behind a single blind. We then presented the dog with two blinds, positioned equidistantly in front of the dog in a central area of the room, and the experimenter placed a single treat behind one of the blinds as the dog watched. Again, the dog was immediately released to search for the hidden food, and the experimenter helped the dog to locate the reward if they did not quickly find it independently. This process was repeated until the dog independently approached the baited blind first on at least two occasions. Across this familiarization stage and both subsequent tasks involving blinds, treats were positioned in the corner of the baited blind(s) such that they were not visible to the dog unless the dog looked down into the blind from above. In the two-location and delayed search tasks that used blinds, baited locations were counterbalanced across trials and followed the same pre-determined order for all dogs.

Two-location task This task measured a dog's ability to remember the location of one or two food rewards hidden behind an array of four blinds that were arranged in a square configuration in front of the dog (1.22 m between blinds; Fig. S1).

Warm-ups. The task began with four warm-up trials in which a treat was hidden behind a single blind. As the dog watched, the experimenter approached and visibly placed a treat behind one blind. Five seconds after placement of this treat, the handler released the dog to search for the reward. The dog had 20 s to search for the hidden treat. If they did not locate it within 20 s, the experimenter showed them the location of the treat and allowed them to eat it, then proceeded to the next warm-up trial. On each trial, we recorded the first location that the dog searched, operationally defined as the dog's snout entering the corner of the blind or the dog's head being positioned above the area behind the blind with a downward head orientation. The reward was hidden behind each of the four blinds one time across the four warm-up trials, following the same pre-determined order for each dog.

Test trials. After the warm-up stage, twelve test trials were conducted. Test trials were identical to familiarization trials with the exception that on half of the test trials, rewards were hidden in two different locations before the dog was allowed to search, thus requiring dogs to maintain a representation of the second item's location in memory while retrieving the first item [40]. When two rewards were hidden, the experimenter placed them sequentially. In all trials, the dog was released to search 5 s after the hiding of the first reward. This 5-s delay served to avoid confounding between the number of baited locations (one or two blinds) with delay (as it required more time to place treats in two locations than one). Five seconds provided sufficient time for the experimenter to bait a second location after the first and return to their starting position to release the dog. Standardizing this duration ensured consistency across trials in which different numbers of rewards were hidden. As in warm-ups, dogs were allowed to search the array until all rewards had been located, or until 20 s elapsed, whichever occurred first. If the dog did not locate all the rewards within 20 s, the remaining rewards were removed from the array prior to the subsequent trial. We recorded the first location searched on trials in which a single reward was hidden, and the first two locations searched on trials in which two rewards were hidden. If the dog moved forward from the start line but did not choose any blinds within 20 s, the trial was marked "no choice" and was not repeated. However, if the dog never left the start line within 20 s, the trial was repeated. If the dog did not leave the start line on two consecutive trials or four trials across the course of the task, the task was discontinued.

The dependent measures for this task were (1) the proportion of correct searches on one-location trials, (2) the proportion of correct first searches on two-location trials, (3) the proportion of correct second searches on two-location trials, and (4) the proportion of two-location trials in which both rewards were recovered without any search errors (i.e., both the first and second locations searched were correct). Dogs who searched on less than half of the administered trials were removed from analysis ($N = 3$).

Delayed search task This task assessed a dog's ability to remember the location of a hidden reward across increasing retention intervals, and in the context of distractions. The procedure was similar to the working memory task described in Bray et al. [20], but with the addition of distractions and some modifications to help ensure continued motivation throughout the task.

Familiarization. First, to familiarize dogs with the procedure that would be used as a distraction in test trials (see below), a trial was performed in which the experimenter walked rapidly towards the dog and placed a treat on the ground in front of the dog, then immediately turned around and returned to their starting position.

Warm-ups. Four blinds were arranged in an array 2.74 m from the dog's starting point, such that all blinds were equidistant from the dog (Fig. S2). On each trial, the experimenter showed the dog a treat, and as the dog watched, placed it behind one of the blinds. After the experimenter returned to their starting position, the dog was immediately released and given 20 s to search for the hidden reward. We recorded the first location that the dog searched, operationally defined as the dog's snout entering the corner of the blind or the dog's head being positioned above the area behind the blind with a downward head orientation. If the first location the dog searched was incorrect, we allowed the dog to continue searching until 20 s had elapsed, or until the reward was located, to prevent declines in motivation associated with failure to obtain a reward. As in the warm-up phase of the two-location task, if the dog did not locate the treat within 20 s, the experimenter led them to the baited location and allowed them to eat the

treat. The reward was hidden once behind each of the blinds across a series of four warm-up trials.

Test trials. The procedure for test trials was identical to that for warm-up trials with the following exceptions. In test trials, dogs were required to wait for a specified duration before being released to search. We used three fixed delays of 10 s (4 trials), 20 s (2 trials), and 40 s (2 trials), which were administered sequentially. The first two test trials were identical to the warm-up trials except that dogs were required to wait 10 s before searching. In the remaining six test trials, we implemented a distraction to prevent the dogs from using non-mnemonic strategies such as sustained attention or body orientation towards the hidden reward [41]. In these distraction trials, during the delay (after placing a treat behind a blind), the experimenter walked towards the dog and delivered a small food reward on the ground in front of the dog, then returned to their starting position. To consume this reward, dogs were required to lower their heads, breaking any possible fixation on the blinds. Additionally, the dogs' view of the blinds became momentarily obscured by the experimenter's body.

As in warm-ups, we recorded the first location searched, and dogs were allowed to continue searching until the reward was located or 20 s had elapsed, but unlike in warm-up trials, if they did not find the food within that time, the experimenter did not lead them to it. The criteria for discontinuing data collection on this task were the same as those used in the two-location task. The dependent measures for this task were the proportion of correct trials within each trial type (i.e., no delay, 10-s delay, 10-s delay with distraction, 20-s delay with distraction, 40-s delay with distraction). For a small subset of dogs who did not make choices on all trials ($N=9$), we imputed missing data using predictive mean matching, implemented in the mice R package [42].

Session 2

Bowl familiarization This brief familiarization was conducted at the beginning of the second session to introduce dogs to the food bowls used in the cylinder task and spatial reversal tasks and confirm that dogs were motivated to obtain food from these bowls. The experimenter first showed the dog a treat, then placed it in a paper bowl (later used during the cylinder task), lowered the paper bowl to the ground, and said "okay," cueing the handler to release the dog to search for the food. The same procedure was repeated with the metal bowl (later used during the spatial reversal task).

Human interaction This task measured a dog's tendency to approach and interact with the experimenter (who varied across testing sites but was in all cases an adult female). The experimenter was positioned against one edge of the testing space. For dogs large enough to rest their chin on the seat of the chair (43 cm seat height), the experimenter sat in a chair, and for smaller dogs, the experimenter sat on a floor cushion. This difference in experimenter position helped equalize dogs' experience as much as possible by ensuring that the experimenter's body, face, and hands were at similar distances from dogs within the proximity zone regardless of dog body size. It also helped avoid the experimenter needing to lean or reach over the heads of small dogs (who may have been overwhelmed by a "looming" person) when attempting to reach and pet them. The handler was positioned in a corner of the testing space opposite the experimenter. When the experimenter sat down, the task began, and the handler released the dog so they could move freely about the room. Anytime the dog came within a marked proximity zone around the experimenter, the experimenter gently petted the dog. The rectangular proximity zone included the experimenter's seat and was 1.22 m × 0.99 m for large dogs, for whom the experimenter sat in a chair, and 1.22 × 0.76 m for small dogs, for whom the experimenter sat on the floor (Fig. S3). These thresholds were selected based on the distance at which the experimenter could easily reach and pet the dog from their seated position. At 15-s intervals, an audio recording played back through headphones prompted the experimenter to say the phrase "what a good dog!". If the dog was not proximate to the experimenter (as defined above) at the time of the prompt, the experimenter patted their leg to beckon the dog while speaking. If the dog was already proximate to the experimenter, the experimenter continued to pet the dog while speaking. Throughout the entire 90-s task duration, the experimenter gazed at the dog, establishing eye contact if the dog looked to the experimenter's face.

From video, we coded the dog's time in proximity to the experimenter and time spent gazing towards the experimenter (operationalized as the dog's head oriented towards the experimenter's head or torso) as dependent measures. We also coded the total time that the camera's view of the dog's gaze was obstructed (in rare events in which this occurred), and converted all measures of gaze time to proportions, subtracting the duration in which the required viewing angle was obstructed from the denominator.

Cylinder This task measured a dog's motor inhibition and response flexibility in the context of changing task demands. The procedure was modified from Bray et al. [20].

Warm-ups. Warm-up trials used an opaque black cylinder (26.5 cm in length) with an opening at one end (diameter 22.5 cm) that was positioned 1.24 m from the dog (Fig. S4). The experimenter acclimated the dog to approaching the cylinder by performing two trials in which they placed a baited paper bowl (height = 25 cm, diameter = 125 cm), in front of the cylinder and allowed the dog to eat from it. Two "leading trials" were then conducted in which the experimenter guided the dog to the open end of the cylinder by allowing them to follow a food lure into the opening of the cylinder. Subsequently, dogs participated in five familiarization trials in which they watched the experimenter place a food reward inside the cylinder, then were released to access the food from the open end of the cylinder (without a lure). The orientation of the open side of the cylinder was consistent for all dogs.

To prevent declines in motivation associated with failure to obtain a reward, if the dog touched the cylinder but did not obtain the food within 20 s, the experimenter guided them to the correct side and allowed them to eat the food from within the cylinder. On all trials, if 20 s elapsed before the dog either successfully retrieved the food or contacted the apparatus, the trial was marked "no response." During this stage of the task, a dog was allowed to access the food after a "no response" (with prompting from the experimenter as needed), but the trial was then repeated until the dog participated, such that all scored responses reflected the dog's independent actions without the experimenter's assistance.

Inhibitory control test trials. These test trials were identical to warm-up trials with the exception that the food reward was placed inside a transparent cylinder, allowing the subject to see the food reward through the front of the apparatus. A successful response (detouring to the open end of the cylinder) therefore required subjects to resist the prepotent response of approaching the visible food directly. Four test trials were performed. We recorded whether dogs touched the exterior of the cylinder with their snout or paw prior to accessing the food inside the cylinder. As before, if a dog touched the cylinder but did not access the food within 20 s, the experimenter guided them to the opening and allowed them to eat the food. If dogs did not access the reward or contact the apparatus within 20 s ("no response"), the trial was repeated without allowing them to access the food. If a second "no response" occurred consecutively, the experimenter placed the bowl in front of the cylinder twice (as in the first stage of warm-ups) to increase the dog's motivation to continue participating. If "no responses" occurred four total times across the inhibitory control and reversal stages, the task was discontinued.

Reversal test trials. Before beginning the reversal stage, the transparent cylinder was rotated 180° such that the treat was now only accessible from the previously closed side. The dog watched as the experimenter placed the treat into the open side of the cylinder, then was released to search. Eight trials were performed. We recorded whether dogs touched the cylinder with their snout or paw prior to accessing the food as well as the first side of the cylinder the dog's snout crossed on their approach (i.e., the open side or the closed side). If "no responses" occurred, the procedure was identical to that for the inhibitory control trials.

The dependent measures for this task were (1) the proportion of trials that dogs obtained the reward without first contacting the exterior of the cylinder in inhibitory control trials and (2) in reversal trials, and (3) the proportion of trials in which the first side approached was correct during reversal trials. Dogs who made responses on two or fewer inhibitory control trials or four or fewer reversal trials were excluded from analysis.

Spatial reversal This task measured a dog's response flexibility, or ability to adapt their behavior to changing circumstances, in the context of a spatial memory task (modified from Osthaus et al. [43]).

Warm-ups. Warm-up trials served to introduce dogs to the apparatus and procedure for navigating around an opaque panel to obtain a reward placed behind it. Dogs were positioned at a start line 1.83 m in front of a freestanding 0.91 m × 0.91 m panel. In the first warm-up trial, the experimenter showed the dog a treat, placed it into a metal bowl (height = 75 cm, diameter = 180 cm), and then placed the bowl on the ground in front of the panel such that it was visible to the dog. The dog was then released and allowed 20 s to obtain the food reward from the bowl. In the second warm-up trial, the experimenter stood centered behind the panel and held a treat above the panel such that it was visible to the dog. The experimenter then placed the treat into the bowl and lowered the bowl behind the panel, audibly placing it on the ground. The experimenter turned around to face away from the dog as the handler released the dog to search for the food behind the panel, which could be obtained by detouring to either the left or right sides of the panel. If a dog did not obtain the reward within 20 s, the trial was repeated, and if the dog failed to obtain the reward within three trials, data collection was discontinued. The side from which the dog approached on this trial determined the setup for the remaining warm-up trials. In all subsequent trials, the procedure was identical with the exception that a barrier (0.61 m wide × 0.91 m tall) was positioned behind the front panel (Fig. S5), which prevented the dog from accessing the food from one side. The barrier blocking access to the food from one side was not visible to the dog from the dog's start line on each trial. For the remainder of warm-ups, the closed side was the side not initially chosen (i.e., if a dog navigated to the left side of the panel in the second warm-up trial, the barrier prevented approaches to the right in subsequent trials). The barrier was constructed of wire mesh such that if the dog approached the closed side, they could see (but not access) the food reward, as well as the open path on the opposite side of the apparatus. During warm-up trials, dogs were allowed up to 20 s from the start of the trial to use the open path to access the reward. The trial ended when the dog accessed the reward. During both the warm-up and test trials, if a dog did not access the food within the allotted trial time limit—either by approaching the closed side but being unable to detour to the open side (failure to solve on their own) or by leaving the start line but not approaching either side (scored as "no choice," which was neither correct nor incorrect and subsequently repeated)—the dog was led to the open side by the handler and allowed to obtain the reward. The handler showing the dog how to access food served to maintain motivation to search and equalize dogs' experience of accessing food at the open side, ensuring that on every trial, if they left the start line, dogs gained experience obtaining the reward via the open path. However, if the dog did not leave the start line within 20 s, the trial was repeated without leading them to the open side. On each trial, we recorded the side first approached, operationally defined as the side on which the dog's snout first crossed the front panel. Before moving on to the next phase, dogs were required to meet a criterion of first approaching the open path in three of four consecutive trials (not including their initial side choice from the second warm-up trial).

Reversal test trials. In each reversal block, the barrier was shifted to the opposite end of the panel, closing the path that was open on the preceding trials, and opening the path on the opposite side. Thus, dogs were required to inhibit the previously successful response and instead employ a new response to access the reward. Trials were repeated in the reversed configuration until the dog's first approach was to the open path in three of four consecutive trials. Upon meeting this criterion, the position of the barrier was reversed again, and another block of trials was administered. This general procedure continued until the dog met the criterion on a total of three reversal blocks, or a total of 30 trials (across blocks) had elapsed during this stage of the task (reversal trials). Throughout all phases of the spatial reversal task, no delay was imposed between trials; rather, the subsequent trial began as soon as the dog was repositioned at the start line.

Based on pilot studies in which several dogs required substantial time to obtain the reward independently during reversals, the maximum trial duration allowed for reversal trials was longer than that for the warm-up trials. The maximum trial time was set to 2 min for the first reversal trial (i.e., the first trial of the first reversal block) and 1 min for all subsequent trials. Besides this change, procedures for reversal trials were identical to warm-up trials. The dependent measures for this task were the total number of reversal blocks completed, the maximum number of trials required to meet criterion on any block, and the

number of unused trials remaining out of 30 possible trials at the conclusion of the task (e.g., 0 for any dog that did not complete three reversals within 30 trials).

Statistical analysis

All analyses were performed using R Statistical Software (v4.3.0) [44]. Inter-rater reliability (comparing scores from a second rater to the primary rater) was assessed using Cohen's kappa for categorical measures and intraclass correlation for continuous measures. Reliability was good to excellent [45, 46] for all measures (Table S1).

To generate summary scores for each task, we used principal component analysis with a correlation matrix for all dependent measures on the task as the data matrix. We retained the first component and used scores on this component as the task score. The rationale for retaining one component per task was to attain a single summary measure of performance for each task, to be weighted equally in the data matrix used for generating overall scores across tasks. To generate the overall score across tasks, we performed principal component analysis with a correlation matrix of the task scores as the data matrix, extracting scores for the first component as an overall measure of performance. All task summary and overall scores were standardized (mean = 0, SD = 1) prior to statistical analysis.

To assess associations between cognitive performance and age, we used Bayesian linear models with fixed effects for dog age, sex, and test location. Age was modeled using a second-order orthogonal polynomial term to capture potential curvilinear relationships between age and cognitive performance. Models were fit using the brms R package [47] which provides an R interface to the Stan programming language [48]. To assess whether the associations between age and cognitive performance differed between dogs of different body sizes (as a proxy for expected lifespan), we conducted secondary analyses of overall battery scores including an interaction between age and a discretized body mass variable with three levels (small: < 15 kg, medium: ≥ 15 < 30 kg, large: ≥ 30 kg). Conditional effects from this model were used to compare associations with age across the different body mass categories and Pareto smoothed importance-sampling leave-one-out cross-validation was used to compare models with and without the age × body mass interaction. All models were fit using four independent chains which were merged for the posterior distribution. We used weakly informative priors for the beta coefficients (normal(mean = 0, SD = 10)), assessed model adequacy via posterior predictive checks, and summarized uncertainty in parameter estimates using 90% credible intervals.

Results

Two-location task

All dependent measures loaded positively on a principal component which accounted for 51% of the variance. The loading for first search accuracy in two-location trials was considerably weaker than the other three loadings (component loadings, first search accuracy, one-location trials = 0.73; first search accuracy, two-location trials = 0.33; second search accuracy, two-location trials = 0.74, errorless two-location recovery = 0.93). In the linear model, coefficients for both the first- and second-order age terms were negative but characterized by substantial uncertainty, with credible intervals including zero in both cases (Fig. 1A; Table 2).

Delayed search task

All five dependent measures loaded positively and substantially on a principal component (component loadings: 0-s delay, no distraction = 0.79; 10-s delay, no distraction = 0.72; 10-s delay, distraction = 0.64; 20-s delay, distraction = 0.55, 40-s delay, distraction = 0.44) which explained 41% of variation across measures. In the linear model, the beta coefficients for the first- and second-order terms for age were both negative, with 90% credible intervals excluding 0 (Fig. 1B, Table 2).

Human interaction

A principal component explained 73% of variance across the dependent measures, which loaded in opposite directions (proximity to experimenter = 0.85, gaze at experimenter = -0.85). For statistical modeling, we used a Yeo-Johnson transformation of the principal component scores, which improved model

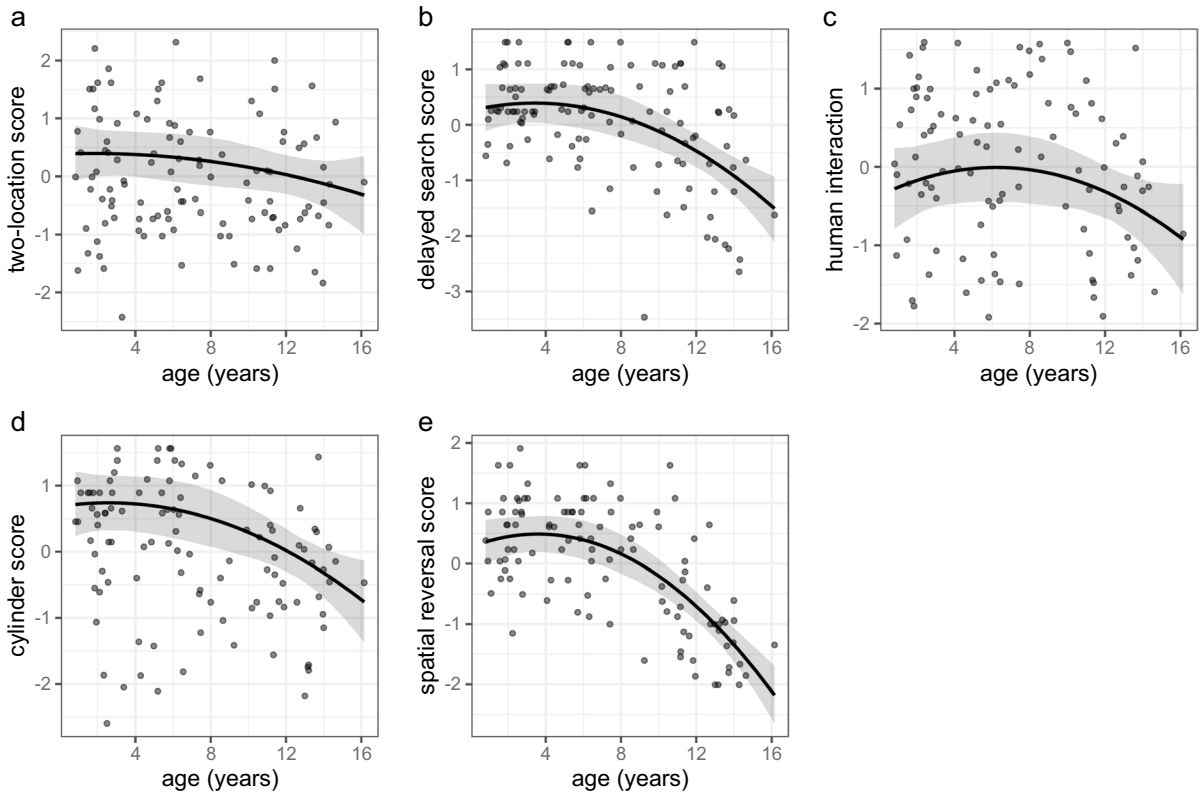

Fig. 1 Associations between age and task performance for the **A** two-location, **B** delayed search, **C** human interaction, **D** cylinder, and **E** spatial reversal tasks. The shaded interval reflects the 90% credible interval from a linear model predicting task performance as a function of a second-order polynomial term for age, with covariates for dog sex and test location. Outcome variables are in standardized units (mean = 0, standard deviation = 1). In panels **A**, **B**, **D**, and **E**, higher scores reflect better performance in the cognitive task. In panel **C**, higher scores reflect more time in proximity to, and less time gazing at the experimenter

adequacy relative to a model fit with the non-transformed data. Coefficient estimates were negative for the first- and second-order age terms, but credible intervals included zero in both cases (Fig. 1C; Table 2).

Cylinder task

The three dependent measures all loaded positively on a principal component which explained 57% of the variance. Loadings were strongest for variables reflecting whether the dog touched the exterior of the transparent cylinder during reversal trials (0.92) and the accuracy of the first side approached during reversal trials (0.85) than for the variable reflecting whether dogs touched the exterior of the cylinder during the pre-reversal phase (0.40). The coefficients for the first- and second-order age terms from the linear model were both negative but with more uncertainty regarding the sign of the second-order term (Fig. 1D; Table 2).

Spatial reversal task

All variables loaded highly on a principal component which explained 85% of the variance. This component had positive loadings for the number of blocks completed (0.89) and number of trials remaining at completion (0.89), and a negative loading for the maximum number of trials required in a block (−0.98). For statistical modeling, we used a Yeo-Johnson transformation of the principal component scores, which improved model adequacy relative to a model fit with the non-transformed data (assessed

Table 2 Results from linear models estimating the association between age and performance on individual tasks

Predictor	Two-location		Delayed search		Human interaction		Cylinder		Spatial reversal	
	Beta	90% CI[1]	Beta	90% CI[1]	Beta	90% CI[1]	Beta	90% CI[1]	Beta	90% CI[1]
Age (polynomial)										
1st degree	−1.9	−3.8, −0.07	−4.6	−6.2, −3.1	−1.1	−2.9, 0.67	−3.8	−5.6, −2.0	−6.3	−7.6, −5.0
2nd degree	−0.56	−2.3, 1.2	−1.9	−3.5, −0.33	−1.4	−3.2, 0.30	−1.3	−3.0, 0.38	−2.7	−4.0, −1.4
Sex										
Male	−0.07	−0.40, 0.27	0.16	−0.13, 0.45	−0.01	−0.35, 0.33	−0.19	−0.51, 0.14	0.05	−0.19, 0.29
Test location										
Canine Companions	−0.27	−0.70, 0.17	0.20	−0.17, 0.58	0.10	−0.36, 0.57	−0.45	−0.90, 0.00	0.35	0.03, 0.68
Sit! Stay! Play!	−0.41	−0.78, −0.04	−0.22	−0.56, 0.11	0.46	0.04, 0.86	−0.58	−0.97, −0.18	−0.18	−0.47, 0.11

[1]*CI* credible interval

using posterior predictive checks). Beta coefficients from the linear model were negative for the first- and second-order age terms, with credible intervals excluding zero in both cases (Fig. 1E; Table 2).

Overall scores

To generate an overall score across tasks, we used principal component analysis using the task-level scores as the data matrix. Scores for all tasks, except for human interaction, loaded strongly and positively on a principal component that explained 38% of variation in task scores (component loadings: two-location = 0.63; delayed search = 0.72; cylinder = 0.60, spatial reversal = 0.77; human interaction = −0.21). To explore the utility of an overall score using data from only the two tasks with the strongest relationship with age (delayed search and spatial reversal), we used a similar approach, generating principal component scores from just these tasks. Both tasks loaded highly on this component (loadings = 0.86), which explained 73% of variation.

A linear model of overall scores derived from the analysis of all five tasks supported negative coefficients for both the linear and second-order age terms (Table 3; Fig. 2A). On average, a 4-year-old dog was estimated to perform ~ 0.7 standard deviations above the sample mean whereas a 14-year-old dog was estimated to perform ~ 0.8 standard deviations below the sample mean. Analysis of overall scores based on only the delayed search and spatial reversal tasks yielded similar results, with negative coefficients for both the linear and second-order age terms (Table 3; Fig. 2B). In this model, on average, a 4-year-old dog was estimated to perform ~ 0.5 standard deviations above the sample mean whereas a 14-year-old dog was estimated to perform ~ 1.3 standard deviations below the sample mean.

Associations between age and performance in relation to expected lifespan

Because expected lifespan differs dramatically between small and large dogs [49], we assessed whether the associations between age and cognitive performance differed between dogs of different body masses. The estimated associations with age were generally similar for small (< 15 kg), medium (15, 30) kg), and large (≥ 30 kg) dogs (Fig. 2C, D). In the

Table 3 Results from linear models estimating the association between age and overall scores, calculated across all five tasks, or only the two tasks with the strongest associations with age

Predictor	Five task score		Two task score	
	Beta	90% CI[1]	Beta	90% CI[1]
Age (polynomial)				
1st degree	−5.6	−7.2, −3.9	−6.5	−7.8, −5.2
2nd degree	−2.1	−3.6, −0.57	−2.6	−4.0, −1.3
Sex				
Male	0.00	−0.30, 0.29	0.11	−0.13, 0.35
Test location				
Canine Companions	−0.04	−0.42, 0.34	0.29	−0.04, 0.62
Sit! Stay! Play!	−0.50	−0.86, −0.14	−0.25	−0.54, 0.05

[1] *CI* credible interval

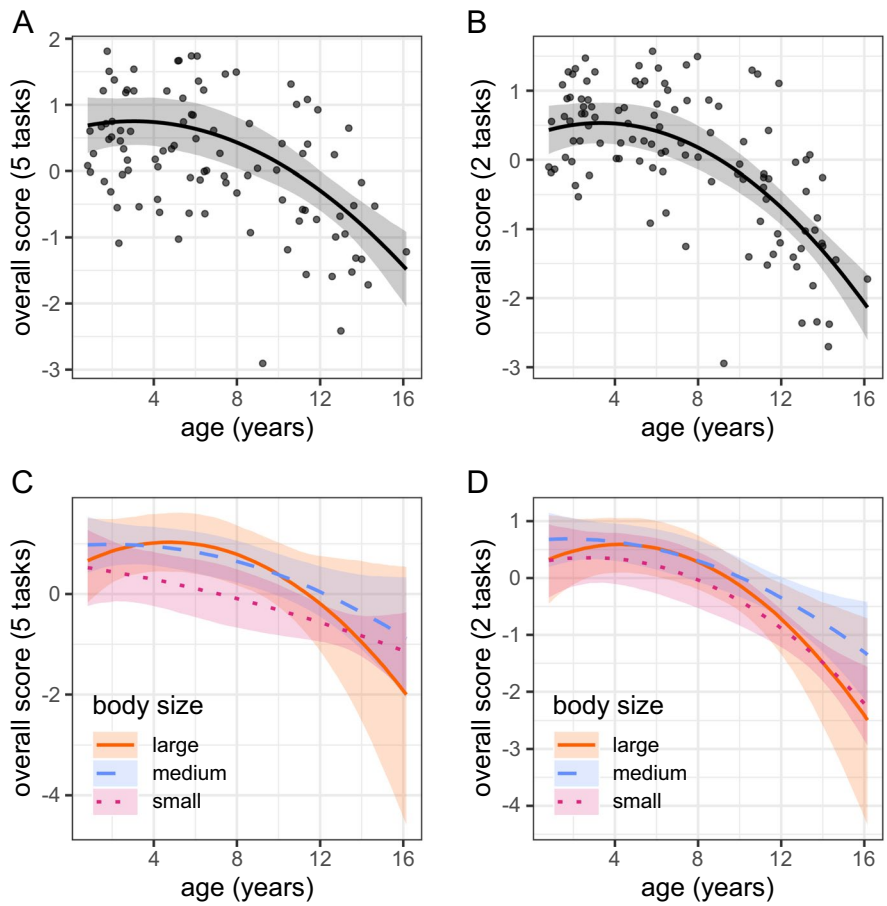

Fig. 2 **A** Associations between age and overall performance based on a summary measure across all five tasks, or **B** the two tasks with the strongest associations with age (delayed search and spatial reversal). **C** Associations with age for small (< 15 kg, $N = 30$), medium (≥ 15 < 30 kg, $N = 62$), and large (≥ 30 kg, $N = 25$) dogs using a summary measure based on all five tasks, or **D** the two tasks with the strongest association with age. In all panels, lines reflect the conditional effects from the statistical model, with shaded regions indicating the 90% credible interval for these effects

analysis of overall scores across the full set of five tasks, the association with age was relatively linear in small dogs, compared to curvilinear effects estimated for medium and large dogs (Fig. 2C). However, we did not observe this phenomenon when considering overall scores derived from the two tasks that had the strongest associations with age (Fig. 2D). Comparison of models with and without the age × body mass interaction (based on Pareto smoothed importance-sampling leave-one-out cross-validation) supported models without the interaction in both cases, although the differences between models were small (difference in expected log pointwise predictive density ± standard error, 5-task model = 2 ± 3.1; 2-task model = 3 ± 2.4).

Experiment 2

In Experiment 2, we assessed the feasibility of using this test battery in clinical studies of canine cognitive dysfunction.

Methods

Thirty-five dogs participating in ongoing clinical studies were tested at the Colorado State University (CSU) College of Veterinary Medicine & Biomedical Sciences (CSU IACUC protocols #4016 & #3384). Demographic information for participating dogs is provided in Table 4.

Cognitive tests were administered by study staff in an exam room during research visits, when clients and their dogs visited the veterinary hospital for medical evaluations. A total of six experimenters administered study tasks. Arizona Canine Cognition Center staff provided initial training for two experimenters at CSU, who subsequently provided training for additional personnel as needed. Arizona Canine Cognition Center staff reviewed CSU study video on an ongoing basis to ensure methodological consistency, and a subset of tests in which experimental error occurred were removed from analyses (see below).

Vital signs and any necessary bloodwork or samples for the clinical study were obtained prior to cognitive testing. Dogs were then taken to the testing area and given time to acclimate to the environment before the tests were administered. Owners were typically absent during testing sessions, but there was a small subset of dogs ($N=4$) where the owner performed the handler role. Research visits were scheduled ~ 6 months apart and dogs typically completed the two-location ($N=35$) and delayed search ($N=35$) tasks during the first visit, and the cylinder ($N=19$), spatial reversal ($N=19$), and human interaction ($N=15$) tasks during their second visit. Because second visits had not yet occurred for many dogs, sample sizes were smaller for the latter tasks. A subset of dogs ($N=9$) completed all tasks during a single initial study visit, with a break between the tasks typically administered in separate sessions. Lastly, a subset of data was excluded from some tasks due to experimental error during task administration (two-location, $N=1$; delayed search, $N=2$; cylinder, N = 2; spatial reversal, $N=7$).

Dog owners in these studies completed the CAnine DEmentia Scale (CADES), a validated questionnaire that was developed to characterize normal aging, and mild, moderate, and severe cognitive dysfunction in dogs [14]. Clients completed the CADES at each study visit and we calculated the average CADES score between the two study visits during which cognitive testing was performed. In the current sample, 19 dogs had mean CADES scores consistent with normal aging, and a smaller number of dogs met criteria for mild ($N=8$), moderate ($N=6$), or severe ($N=2$) cognitive dysfunction.

To rule out other potential causes for behavioral changes indicative of cognitive dysfunction, all dogs underwent a series of tests and evaluations. These included a complete blood count (CBC), chemistry panel, thyroid panel, urinalysis, thoracic radiographs, abdominal ultrasound, brain MRI, and spinal tap to rule out infectious or inflammatory diseases of the central nervous system.

Table 4 Selected characteristics of companion dog participants in Experiment 2. *IQR* interquartile range

	$N=35^1$
Sex	
Intact male	1 (2.86%)
Neutered male	18 (51.43%)
Spayed female	16 (45.71%)
Breed category	
Mixed breed	10 (28.57%)
Purebred	25 (71.43%)
Cognitive Dysfunction Severity (CADES)	
Normal	19 (54.29%)
Mild	8 (22.86%)
Moderate	6 (17.14%)
Severe	2 (5.71%)
Weight (kg)	24.50 (18.00, 31.25)
Age (years)	9.27 (4.90, 12.02)

[1] *n* (%); median (IQR)

Results

As in Experiment 1, inter-rater agreement was good to excellent for all measures (Table S2). To assess feasibility, we considered the percentage of dogs with sufficiently complete data for calculation of task summary scores. For the spatial reversal, cylinder task, and human interaction tasks, all dogs tested

had sufficiently complete data, demonstrating high feasibility. For the delayed search task, one dog did not choose on enough trials for calculation of a summary score, resulting in 97% data completeness, and the CADES score for the sole subject with insufficient data was consistent with normal aging. Lastly, in the two-location task, 7 dogs did not have sufficient data for calculation of summary scores (CADES classifications: normal aging, $N=3$; mild impairment, $N=1$; moderate impairment, $N=2$; severe impairment, $N=1$). This was often the result of dogs only choosing a single location during the 20-s search period even on trials where two treats were hidden, which precluded the calculation of second choice accuracy. The percentage of dogs with sufficient data for analysis on this task was considerably lower in Experiment 2 (80%) than in Experiment 1 (93%), suggesting that this task may have the lowest feasibility in clinical settings.

Summary scores for each task were generated using principal component models from the dataset in Experiment 1. Due to the relatively small sample in this feasibility study, and limited representation of dogs with greater than mild cognitive dysfunction, we did not formally model associations between CADES scores and performance on the cognitive tasks. The distributions of scores on each task, in relation to categories of impairment based on the CADES, are shown in Fig. 3. Among normal agers and mildly impaired dogs, task performance and variability was generally comparable to the standard normal distribution (mean = 0, SD = 1) generated from dogs in Experiment 1, with the exception of the spatial reversal task, for which mean scores were higher in

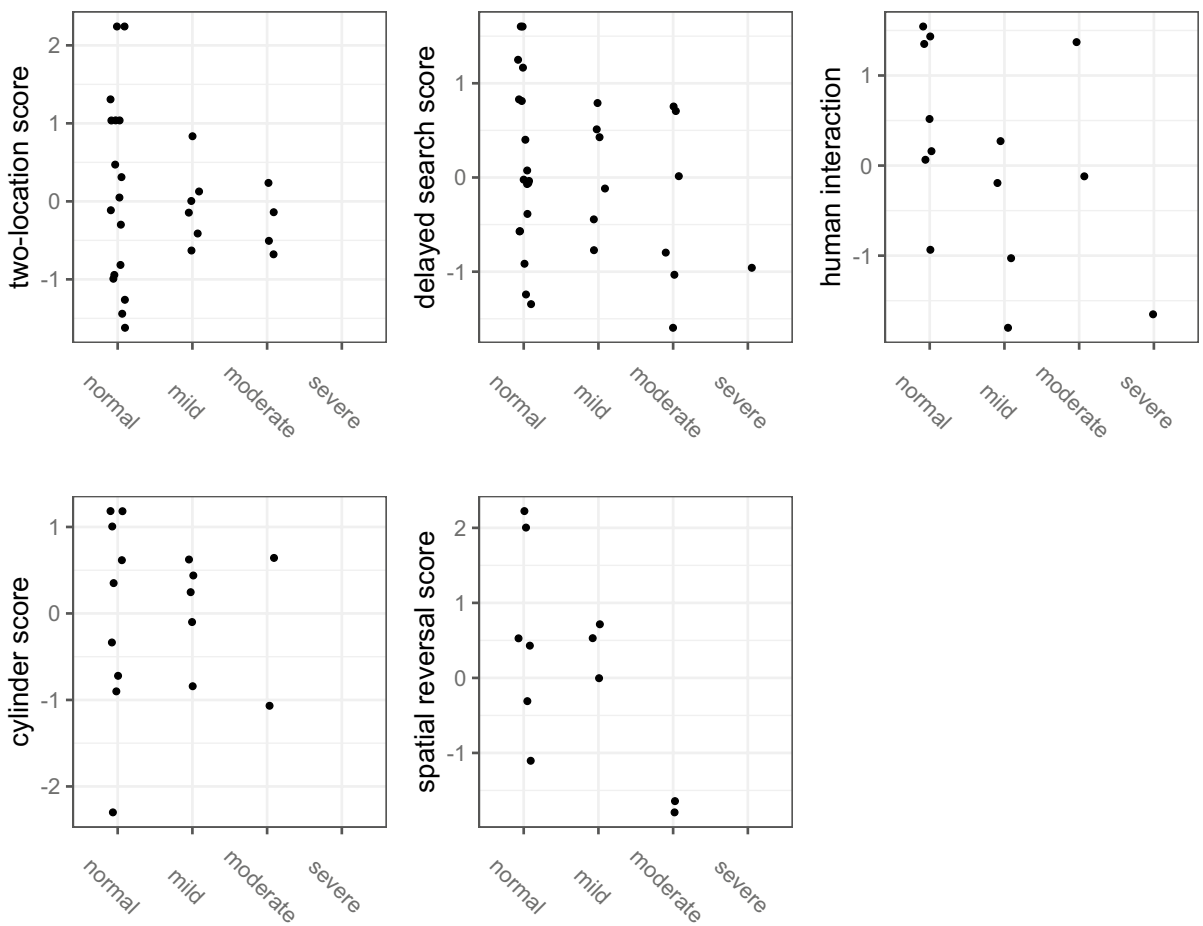

Fig. 3 Performance of dogs in Experiment 2 (clinical setting) as a function of the extent of clinical symptoms of dementia as measured by the CAnine DEmentia Scale (CADES)

Experiment 2 (Experiment 2 mean ± SD scores for normal aging and mildly impaired dogs: two-location task = 0.02 ± 1.07; delayed search task = 0.05 ± 0.79; human interaction = 0.11 ± 1.04; cylinder task = −0.01 ± 0.94; spatial reversal = 0.59 ± 1.1). However, this deviation in mean scores on the spatial reversal task should be interpreted cautiously given the small sample for this task in Experiment 2 ($n = 11$). Dogs with moderate or severe cognitive impairment tended to score within the range of animals classified as normal to mildly impaired, although the mean performance in the moderate and severely impaired groups was somewhat lower in all cases (mean scores among moderately or severely impaired dogs: two-location task = −0.25; delayed search task = −0.4; human interaction = −0.16; cylinder task = −0.54; spatial reversal = −1.71).

Discussion

We developed and implemented a novel battery of five tasks to characterize cognitive aging in dogs. Performance on all tasks was estimated to be negatively associated with age, although the strength of these relationships varied across tasks. Older dogs performed at lower levels on two measures of spatial memory, with the strongest association for the delayed search task. Older dogs also scored lower on the two executive function tasks requiring subjects to inhibit prepotent responses or to exhibit response flexibility in a spatial reversal task. Lastly, in a simple assessment of social interaction with an experimenter, older dogs spent less time in proximity to, but gazed more, at the experimenter, although there was considerable uncertainty in the beta coefficients for age in the task. Overall, the strongest associations with age were observed in the delayed search and spatial reversal tasks, suggesting that these measures may be particularly useful in future studies of dog cognitive aging. Our results for the memory and executive function tasks align with those of previous studies demonstrating impairment in older dogs on visuospatial memory tasks [26, 50] as well as on reversal learning and spatial detour tasks designed to measure executive function [28, 30, 31, 51]. However, our results from the social interaction task differ from those of other studies, which have reported decreased gaze at an experimenter in older dogs. These differences may be accounted for by the context in which gaze was measured. Specifically, in the current studies, we measured gaze in a naturalistic social interaction during which the experimenter spoke to, gazed at, and (if approached) petted the dog. Previous studies reporting decreased gaze duration in older dogs have employed gaze as a measure of sustained attention often in situations when the dog's attention is initially captured by a food reward or an unusual behavioral act performed by the experimenter [52, 53].

With the exception of the social measure, summary measures for all tasks loaded positively on a principal component, from which we extracted scores as an overall measure of performance. These overall scores were strongly associated with dog age with negative estimates for both the first- and second-order age terms in the linear model. On average, a 14-year-old dog was estimated to score ~ 1.5 standard deviations lower than a 4-year-old dog. A secondary analysis using summary scores from only the delayed search and spatial reversal task largely recapitulated this effect, but with more pronounced differences between middle-aged and senior dogs. Consistent with previous studies of cognitive aging in laboratory populations of dogs [6], some degree of age-related impairment was detectable beginning in middle age (~ 5–9 years). Although we found robust associations between age and performance, we also tested relatively few very old dogs. Given these characteristics of our sample, it is possible that among the oldest dogs, cognition is impaired to an extent beyond that which we were able to capture in the current work.

Given large differences in expected lifespan for dogs of different body masses, we assessed whether associations between performance and age differed between small, medium, and large dogs. We found no strong evidence for associations with age being moderated by body mass, consistent with other studies of cognitive aging in companion dogs [27, 30], but see Turcsán and Kubinyi [54]; however, our sample had relatively few extremely large or small dogs, potentially limiting our ability to detect such an effect. Future longitudinal studies will be required to determine whether the timing and/or rate of cognitive aging covaries with body mass.

In Experiment 2, we assessed the feasibility of implementing these tasks in clinical settings and explored potential associations with clinical surveys used to quantify cognitive impairments in dogs. On

average, dogs categorized as having moderate or severe impairments (based on the CADES) scored lower on most tasks, but the range of scores for these dogs overlapped with dogs characterized as having mild cognitive impairment or as normal cognitive agers. In the current sample, relatively few dogs ($n = 8$) met criteria for moderate or severe cognitive impairment and future work in larger clinical studies will be required to robustly assess concordance between experimentally derived measures of cognitive function and owner report surveys. The current studies suggest that the simple measures of memory, executive function, and social behavior from this battery are feasible for use in at least some clinical settings and may provide relatively easily administered objective assessments of cognitive function in aging dogs.

The battery of tests presented here has several advantages relative to other current or historically used approaches. Our assessments of spatial memory and executive function are based on spontaneous problem-solving rather than responses trained via operant procedures. This approach eliminates the lengthy training protocols required to shape behavioral responses and enables collection of key outcome measures in a single testing session. Because individual tasks can be implemented relatively quickly, it is also possible to collect data using multiple tasks within the same testing session, providing opportunities to rapidly characterize different aspects of cognitive function. Lastly, by reducing the needs for specialized testing equipment, these tests can potentially be administered in a wide range of settings, enabling large-scale studies of companion dogs.

Despite these advantages, these tests are subject to several important limitations. First, most tasks involve participation for food rewards and cannot be meaningfully administered if a subject is not sufficiently motivated to attain these rewards (or has major mobility impairments that preclude participation in physical search tasks). Food motivation can vary considerably between individuals, within individuals across contexts, and may be reduced in novel settings such as visits to research labs or veterinary clinics. Although we had relatively high completion rates, there was also potential for selection bias (e.g., dog owners with food-motivated dogs who would be comfortable in the testing environment potentially being more likely to participate). In the current work, tests were successfully administered in four different settings including a university research center, a dog daycare, a service dog training center, and a veterinary hospital, highlighting diverse environments in which such studies are possible. Nonetheless, some dogs could not be tested successfully due to lack of food motivation or nervousness about the environment or unfamiliar experimenters. For the success of large-scale projects using these or similar measures, it may be important to incorporate inclusion criteria related to food motivation and to identify additional testing environments or acclimation procedures that minimize attrition. Additionally, although tasks were successfully implemented in diverse settings, we observed variation in performance on some tasks, and in overall scores across tasks, associated with the testing location. In the current design, we cannot determine whether these effects stem from variation in the physical testing environments, population differences between dogs tested at the different sites, or other unidentified sources of variation.

Second, only one dog in Experiment 2 had a CADES score consistent with severe cognitive dysfunction and the feasibility of these tests with severely impaired dogs remains largely unexplored. Therefore, assessing both the feasibility and the utility of these measures in objectively capturing cognitive impairments associated with advanced stages of Canine Cognitive Dysfunction remains an important future research priority. Whereas experimental measures of cognitive function will be critical for understanding all stages of Canine Cognitive Dysfunction, we expect that the current tests may be most useful for characterizing initial, mild cognitive impairments, that precede severe manifestations of dementia. Given that early intervention is widely believed to present the best opportunity for slowing or reversing the course of disease [55], the development of validated assays to capture the early stages of dementia remains a high priority. Current diagnostic tools in dogs rely on owner report of behavioral symptoms that may only arise in later stages of dementia; therefore, cognitive assays, such as those presented here, have the potential to fill important needs in both research and clinical contexts.

Third, the present data are cross-sectional, and while we identified robust associations between task performance and age, longitudinal studies using these approaches will be essential in future

research. Determining test–retest reliability, quantifying potential practice effects across different testing intervals, and determining whether within-individual changes can predict the development of dementia are essential next steps that can only be completed in the context of longitudinal designs.

Fourth, as our tasks were designed to maintain high levels of motivation, dogs were allowed to obtain food rewards after incorrect initial searches, and under some conditions, failure to independently access the food within a time limit. It is possible that obtaining a food reward despite failure to solve the problem in the intended way prevented some dogs from learning the "goals" of the task. However, we do not believe this to be the case generally. If dogs were motivated to obtain the reward (as evidenced by the very high rates of participation), we expected that they would minimize energetic costs and maximize behavioral efficiency when attempting to obtain the reward. Nonetheless, the relatively low cost of errors remains an important consideration. This is particularly relevant for the two-location task, in which successful performance often required dogs to bypass nearby and easily accessible blinds when navigating to more distant, baited locations. We observed that dogs sometimes had difficulty bypassing the nearer, non-baited blinds, possibly due to task demands involving inhibitory control, but also likely due in part to the low cost of inspecting the proximal non-baited locations. These factors may partially explain why we observed weaker age effects on this task, as both young and old dogs had somewhat poor performance.

Lastly, although the measures presented here can be implemented much more rapidly than conventional operant tasks, the current design still requires two ~ 30–60-min testing sessions. Development of a more streamlined assessment that could be implemented in a single brief session would have major advantages for facilitating large-scale studies with companion dogs. Based on the current results, the delayed search and spatial reversal tasks would be good candidates for a reduced set of measures that (1) have strong associations with age, (2) assess cognitive processes implicated in both canine cognitive dysfunction and Alzheimer's disease, and (3) could be paired in a single short assessment suitable for implementation in diverse settings.

A streamlined assessment that can objectively measure canine cognitive aging would complement large-scale studies with companion dogs [such as the Dog Aging Project, 3] and could serve as an important tool in both Alzheimer's disease research and veterinary medicine. The development and validation of these approaches has the potential to be highly valuable for understanding the cognitive correlates of candidate dementia biomarkers, assessing the functional impacts of treatments and interventions designed to prolong cognitive health, and contributing to basic research that addresses the causes and consequences of cognitive aging.

Acknowledgements We thank Canine Companions and Sit! Stay! Play! for facilitating and supporting research at their facilities. We thank Chloe Burkholder, Elizabeth Carranza, Grace Davis, Emily Davis, Abby Flyer, Catherine Fry, Gabi Garcia, Ana Gonzales, Emily Hilyard, Emma Kristal, Kevin McLaughlin, Belinda Morton, Tyler Rodriquez, Alexa Sell, Caitlyn Seymour, Olivea Spencer, Brianna Tamayo, Natya Thirunagari, Christina Cawood, Alessandra Ostheimer, Angelina Schirle, Rachel Shaberman, Olivia Gittings, and Nova Vogel for their help with research logistics, data collection, and/or video scoring. Lastly, we thank the many individuals and dogs in our community who participated in these studies.

Dog Aging Project consortium authors Joshua M. Akey[1], Brooke Benton[2], Elhanan Borenstein[3,4,5], Marta G. Castelhano[6,7], Amanda E. Coleman[8], Kate E. Creevy[9], Kyle Crowder[10,11], Matthew D. Dunbar[11], Virginia R. Fajt[12], Annette L. Fitzpatrick[13,14,15], Unity Jeffery[16], Erica C. Jonlin[2,17], Matt Kaeberlein[2], Elinor K. Karlsson[18,19], Kathleen F. Kerr[20], Jonathan M. Levine[9], Jing Ma[21], Robyn L. McClelland[20], Daniel E.L. Promislow[2,22,23], Audrey Ruple[24], Stephen M. Schwartz[14,25], Sandi Shrager[26], Noah Snyder-Mackler[27,28,29], M. Katherine Tolbert[9], Silvan R. Urfer[2], Benjamin S. Wilfond[30,31]

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA.

[2]Department of Laboratory Medicine and Pathology, University of Washington School of Medicine, Seattle, WA, USA.

[3]Department of Clinical Microbiology and Immunology, Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel.

[4]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel.

[5]Santa Fe Institute, Tel Aviv University, Santa Fe, NM, USA.

[6]Cornell Veterinary Biobank, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA.

[7]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA.

[8]Department of Small Animal Medicine and Surgery, College of Veterinary Medicine, University of Georgia, Athens, GA, USA.

[9]Department of Small Animal Clinical Sciences, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA.
[10]Department of Sociology, University of Washington, Seattle, WA, USA.
[11]Center for Studies in Demography and Ecology, University of Washington, Seattle, WA, USA.
[12]Department of Veterinary Physiology and Pharmacology, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA.
[13]Department of Family Medicine, University of Washington, Seattle, WA, USA.
[14]Department of Epidemiology, University of Washington, Seattle, WA, USA.
[15]Department of Global Health, University of Washington, Seattle, WA, USA.
[16]Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA.
[17]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA.
[18]Bioinformatics and Integrative Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA.
[19]Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[20]Department of Biostatistics, University of Washington, Seattle, WA, USA.
[21]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
[22]Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA.
[23]Department of Biology, University of Washington, Seattle, WA, USA.
[24]Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA, USA.
[25]Epidemiology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
[26]Collaborative Health Studies Coordinating Center, Department of Biostatistics, University of Washington, Seattle, WA, USA.
[27]School of Life Sciences, Arizona State University, Tempe, AZ, USA.
[28]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA.
[29]School for Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA.
[30]Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA, USA.
[31]Department of Pediatrics, Division of Bioethics and Palliative Care, University of Washington School of Medicine, Seattle, WA, USA.

Funding This research was supported by the National Institute on Aging (U19AG057377 and R24AG073137) and the Dogtopia Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Hoffman JM, Creevy KE, Franks A, et al. The companion dog as a model for human aging and mortality. Aging Cell. 2018;17:e12737. https://doi.org/10.1111/acel.12737.
2. Ruple A, MacLean E, Snyder-Mackler N, et al. Dog models of aging. Annu Rev Anim Biosci. 2022;10:419–39. https://doi.org/10.1146/annurev-animal-051021-080937.
3. Creevy KE, Akey JM, Kaeberlein M, et al. An open science study of ageing in companion dogs. Nature. 2022;602:51–7. https://doi.org/10.1038/s41586-021-04282-9.
4. King A. The search for better animal models of Alzheimer's disease. Nature. 2018;559:S13–5. https://doi.org/10.1038/d41586-018-05722-9.
5. Head E. Neurobiology of the aging dog. Age. 2011;33:485–96. https://doi.org/10.1007/s11357-010-9183-3.
6. Head E. A canine model of human aging and Alzheimer's disease. Biochim Biophys Acta BBA - Mol Basis Dis. 2013;1832:1384–9. https://doi.org/10.1016/j.bbadis.2013.03.016.
7. Milgram NW, Head E, Zicker SC, et al. Long-term treatment with antioxidants and a program of behavioral enrichment reduces age-dependent impairment in discrimination and reversal learning in beagle dogs. Exp Gerontol. 2004;39:753–65. https://doi.org/10.1016/j.exger.2004.01.007.
8. Packer RMA, McGreevy PD, Salvin HE, et al. Cognitive dysfunction in naturally occurring canine idiopathic epilepsy. PLoS One. 2018;13:e0192182. https://doi.org/10.1371/journal.pone.0192182.
9. Ruehl WW, Bruyette DS, DePaoli A, et al. Canine cognitive dysfunction as a model for human age-related cognitive decline, dementia and Alzheimer's disease: clinical presentation, cognitive testing, pathology and response to 1-deprenyl therapy. Prog Brain Res. 1995;106:217–25. https://doi.org/10.1016/s0079-6123(08)61218-2.
10. Urfer S, Darvas M, Keene D, et al. Amyloid beta-42 levels in companion dog brains correlate with age and cognitive function. Innov Aging. 2020;4:887. https://doi.org/10.1093/geroni/igaa057.3273.
11. Russell MJ, Bobik M, White RG, et al. Age-specific onset of β-amyloid in Beagle brains. Neurobiol Aging. 1996;17:269–73. https://doi.org/10.1016/0197-4580(95)02072-1.
12. Head E, McCleary R, Hahn FF, et al. Region-specific age at onset of β-amyloid in dogs. Neurobiol Aging. 2000;21:89–96. https://doi.org/10.1016/S0197-4580(00)00093-2.

13. Salvin HE, McGreevy PD, Sachdev PS, Valenzuela MJ. Under diagnosis of canine cognitive dysfunction: a cross-sectional survey of older companion dogs. Vet J. 2010;184:277–81. https://doi.org/10.1016/j.tvjl.2009.11.007.
14. Madari A, Farbakova J, Katina S, et al. Assessment of severity and progression of canine cognitive dysfunction syndrome using the CAnine DEmentia Scale (CADES). Appl Anim Behav Sci. 2015;171:138–45.
15. Salvin HE, McGreevy PD, Sachdev PS, Valenzuela MJ. The canine cognitive dysfunction rating scale (CCDR): a data-driven and ecologically relevant assessment tool. Vet J. 2011;188:331–6. https://doi.org/10.1016/j.tvjl.2010.05.014.
16. Garcia-Ptacek S, Eriksdotter M, Jelic V, et al. Subjective cognitive impairment: towards early identification of Alzheimer disease. Neurol Barc Spain. 2016;31:562–71. https://doi.org/10.1016/j.nrl.2013.02.007.
17. Gauthier SG. Alzheimer's disease: the benefits of early treatment. Eur J Neurol. 2005;12(Suppl 3):11–6. https://doi.org/10.1111/j.1468-1331.2005.01322.x.
18. Salmon DP, Bondi MW. Neuropsychological assessment of dementia. Annu Rev Psychol. 2009;60:257–82. https://doi.org/10.1146/annurev.psych.57.102904.190024.
19. Bensky MK, Gosling SD, Sinn DL, et al. Chapter Five - The world from a dog's point of view: a review and synthesis of dog cognition research. In: Brockmann HJ, Roper TJ, Naguib M, et al., editors. Advances in the study of behavior. Academic Press; 2013. p. 209–406.
20. Bray EE, Gruen ME, Gnanadesikan GE, et al. Dog cognitive development: a longitudinal study across the first 2 years of life. Anim Cogn. 2021;24:311–28. https://doi.org/10.1007/s10071-020-01443-7.
21. Horschler DJ, Hare B, Call J, et al. Absolute brain size predicts dog breed differences in executive function. Anim Cogn. 2019;22:187–98. https://doi.org/10.1007/s10071-018-01234-1.
22. MacLean EL, Herrmann E, Suchindran S, Hare B. Individual differences in cooperative communicative skills are more similar between dogs and humans than chimpanzees. Anim Behav. 2017;126:41–51. https://doi.org/10.1016/j.anbehav.2017.01.005.
23. Shaw RC, Schmelz M. Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. Anim Cogn. 2017;20:1003–18. https://doi.org/10.1007/s10071-017-1135-1.
24. Chapagain D, Range F, Huber L, Virányi Z. Cognitive aging in dogs. Gerontology. 2017;64:165–71. https://doi.org/10.1159/000481621.
25. Mondino A, Khan M, Case B, et al. Winning the race with aging: age-related changes in gait speed and its association with cognitive performance in dogs. Front Vet Sci. 2023;10. https://doi.org/10.3389/fvets.2023.1150590
26. Piotti P, Piseddu A, Aguzzoli E, et al. Two valid and reliable tests for monitoring age-related memory performance and neophobia differences in dogs. Sci Rep. 2022;12. https://doi.org/10.1038/s41598-022-19918-7.
27. Watowich MM, MacLean EL, Hare B, et al. Age influences domestic dog cognitive performance independent of average breed lifespan. Anim Cogn. 2020;23:795–805. https://doi.org/10.1007/s10071-020-01385-0.
28. Piotti P, Szabó D, Bognár Z, et al. Effect of age on discrimination learning, reversal learning, and cognitive bias in family dogs. Learn Behav. 2018;46:537–53. https://doi.org/10.3758/s13420-018-0357-7.
29. Mongillo P, Araujo JA, Pitteri E, et al. Spatial reversal learning is impaired by age in pet dogs. Age. 2013;35:2273–82. https://doi.org/10.1007/s11357-013-9524-0.
30. Van Bourg J, Gunter LM, Wynne CDL. A rapid serial reversal learning assessment for age-related cognitive deficits in pet dogs. Behav Processes. 2021;186:104375. https://doi.org/10.1016/j.beproc.2021.104375.
31. Chapagain D, Wallis LJ, Range F, et al. Behavioural and cognitive changes in aged pet dogs: No effects of an enriched diet and lifelong training. PLoS ONE. 2020;15:e0238517. https://doi.org/10.1371/journal.pone.0238517.
32. Alexander GE, Ryan L, Bowers D, et al. Characterizing cognitive aging in humans with links to animal models. Front Aging Neurosci. 2012;4:21. https://doi.org/10.3389/fnagi.2012.00021.
33. Bäckman L, Jones S, Berger A-K, et al. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. Neuropsychology. 2005;19:520–31. https://doi.org/10.1037/0894-4105.19.4.520.
34. Baudic S, Barba GD, Thibaudet MC, et al. Executive function deficits in early Alzheimer's disease and their relations with episodic memory. Arch Clin Neuropsychol Off J Natl Acad Neuropsychol. 2006;21:15–21. https://doi.org/10.1016/j.acn.2005.07.002.
35. Bediou B, Ryff I, Mercier B, et al. Impaired social cognition in mild Alzheimer disease. J Geriatr Psychiatry Neurol. 2009;22:130–40. https://doi.org/10.1177/0891988709332939.
36. Crawford TJ, Higham S, Mayes J, et al. The role of working memory and attentional disengagement on inhibitory control: effects of aging and Alzheimer's disease. Age. 2013;35:1637–50. https://doi.org/10.1007/s11357-012-9466-y.
37. Freedman M, Oscar-Berman M. Spatial and visual learning deficits in Alzheimer's and Parkinson's disease. Brain Cogn. 1989;11:114–26. https://doi.org/10.1016/0278-2626(89)90009-2.
38. Lithfous S, Dufour A, Després O. Spatial navigation in normal aging and the prodromal stage of Alzheimer's disease: insights from imaging and behavioral studies. Ageing Res Rev. 2013;12:201–13. https://doi.org/10.1016/j.arr.2012.04.007.
39. Verdon C-M, Fossati P, Verny M, et al. Social cognition: an early impairment in dementia of the Alzheimer type. Alzheimer Dis Assoc Disord. 2007;21:25–30. https://doi.org/10.1097/WAD.0b013e318032487a.
40. Sluzenski J, Newcombe NS, Satlow E. Knowing where things are in the second year of life: implications for hippocampal development. J Cogn Neurosci. 2004;16:1443–51. https://doi.org/10.1162/0898929042304804.
41. Krichbaum S, Smith JG, Lazarowski L, Katz JS. Controlling for dogs' (Canis familiaris) use of nonmnemonic strategies in a spatial working memory task. J Exp

Psychol Anim Learn Cogn. 2021;47:364–70. https://doi.org/10.1037/xan0000293.
42. Van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45:1–67.
43. Osthaus B, Marlow D, Ducat P. Minding the gap: spatial perseveration error in dogs. Anim Cogn. 2010;13:881–5.
44. R Core Team. R: A language and environment for statistical computing. 2023. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
45. Altman DG. Practical statistics for medical research. CRC Press; 1990.
46. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
47. Bürkner P-C. brms: an R package for Bayesian multilevel models using Stan. J Stat Softw. 2017;80:1–28.
48. Gelman A, Lee D, Guo J. Stan: a probabilistic programming language for Bayesian inference and optimization. J Educ Behav Stat. 2015;40:530–43.
49. Kraus C, Pavard S, Promislow DE. The size–life span trade-off decomposed: why large dogs die young. Am Nat. 2013;181:492–505.
50. Studzinski CM, Christie L-A, Araujo JA, et al. Visuospatial function in the beagle dog: an early marker of cognitive decline in a model of human aging and dementia. Neurobiol Learn Mem. 2006;86:197–204. https://doi.org/10.1016/j.nlm.2006.02.005.
51. Tapp PD, Siwak CT, Estrada J, et al. Size and reversal learning in the beagle dog as a measure of executive function and inhibitory control in aging. Learn Mem. 2003;10:64–73. https://doi.org/10.1101/lm.54403.
52. Chapagain D, Virányi Z, Wallis LJ, et al. Aging of attentiveness in border collies and other pet dog breeds: the protective benefits of lifelong training. Front Aging Neurosci. 2017;9:100. https://doi.org/10.3389/fnagi.2017.00100.
53. Hoel JA, Templeton GB, Fefer G, et al. Sustained gaze is a reliable in-home test of attention for aging pet dogs. Front Vet Sci. 2021;8:819135. https://doi.org/10.3389/fvets.2021.819135.
54. Turcsán B, Kubinyi E. Differential behavioral aging trajectories according to body size, expected lifespan, and head shape in dogs. GeroScience. 2023. https://doi.org/10.1007/s11357-023-00945-9.
55. Fan D-Y, Wang Y-J. Early intervention in Alzheimer's disease: how early is early enough? Neurosci Bull. 2020;36:195–7.