



# Common marmosets (*Callithrix jacchus*) excel in a one-trial spatial memory test, yet perform poorly in a classical memory task

Sandro Sehner<sup>1,2</sup> · Flávia Mobili<sup>3</sup> · Erik P. Willems<sup>3</sup> · Judith M. Burkart<sup>3,4</sup>

Received: 12 July 2024 / Revised: 26 February 2025 / Accepted: 27 February 2025  
© The Author(s) 2025

## Abstract

When quantifying animal cognition, memory represents one of the most tested domains and is key to understanding cognitive evolution. Memory tests thus play an important role in comparative cognitive research, yet slight variations in the experimental settings can substantially change the outcome, questioning whether different memory tests tap into different memory systems or whether they test memory at all. Here, we first assessed memory performance of 16 common marmosets (*Callithrix jacchus*) in two distinct paradigms varying in their format and delay. First, we examined marmoset memory in a 24-h delay memory test (24 h-DMT) in which they could freely explore an environment with three novel objects of which one contained food. We examined their retention the day after, and the procedure was iterated cumulatively with previous objects remaining in the enclosure until the marmosets had to choose the correct out of 30 objects. Second, we administered a classical delayed response test (DRT) in the same animals with three objects and a maximum delay of 30 s. In the DRT, marmoset performance was poor and not better than chance after 15 s already. However, individuals excelled in the 24 h-DMT, performing above chance level after 24 h even with tenfold the number of objects to choose from compared to the DRT. Moreover, individual performances in the two tests were not correlated, and typical age effects on memory could not be detected in both experiments. Together, these results suggest that the two tests explore different domains, and that the 24 h-DMT examines long-term memory. The outcome of the DRT is more difficult to assign to memory since individuals performed only moderately even in the 0-s delay condition. This puts into question whether this task design indeed tests memory or other cognitive processes.

**Keywords** Common marmosets · Short-term memory · Task format · Physical cognition · Age effects · Long-term memory

## Introduction

Animal cognition is often tested using standardized cognitive tests. For instance, the primate test battery (Banerjee et al. 2009; Herrmann et al. 2007) uses a set of various standardized tests to examine cognitive abilities across domains and species. One of these abilities is memory, which has been assessed in a variety of species. Memory tasks can vary in the way that they are performed, e.g. with regard to the number of choices, or the delays (a few seconds to a couple of minutes) imposed before an animal is allowed to make a decision. Overall, individuals perform worse with an increasing delay in contexts where the maximum delay is only a couple of minutes (Amici et al. 2010; Darusman et al. 2014; Flaim and Blaisdell 2023; ManyPrimates et al. 2022; Reynoso-Cruz et al. 2021; Schubiger et al. 2016). Larger-brained primates such as apes usually outperform the smaller-brained ones and in particular do so

Sandro Sehner and Flávia Mobili have contributed equally to this work.

✉ Sandro Sehner  
SSehner@dpz.eu

✉ Judith M. Burkart  
judith.burkart@aim.uzh.ch

<sup>1</sup> Behavioral Ecology and Sociobiology Unit, German Primate Center, Goettingen, Germany

<sup>2</sup> Centre de Primatologie, Université de Strasbourg, Strasbourg, France

<sup>3</sup> Department of Evolutionary Anthropology, University of Zurich, Zurich, Switzerland

<sup>4</sup> Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland

in longer delay conditions (Amici et al. 2010; van Schaik et al. 2024). However, even if an animal does not make the correct choice, the information about a location of a hidden food reward is not necessarily always lost but merely may be temporarily inaccessible (Martin-Ordas and Call 2011). For instance, chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*) and orangutans (*Pongo pygmaeus*) showed a performance decline from a delay of two minutes up to four hours. However, afterward, performance increased again, and with a delay of 24 h, it was on par again with the delay of two minutes. Overall, performances in memory tests can be influenced by a variety of intrinsic and extrinsic factors. In the following, we will explore three factors that may influence performance in memory tasks: test format, cognitive decline associated with natural aging, and attention allocated to the cognitive test.

First, the format of the test can affect an individual's performance. The oldest and most common way to test memory in animals is through the classical delayed response task (DRT) (Hunter 1913). The experimenter hides food in one of several cups in a way that the animal can see the hiding process and the individual has to choose one of the cups after a predefined delay. Only when the individual chooses the cup where the food is hidden will it receive the reward. This test has been widely applied across primates (e.g., Call 2001; Barth and Call 2006; Amici et al. 2010; Martin-Ordas and Call 2011; Schubiger et al. 2016). Although this paradigm minimizes extraneous variables, thus fostering comparability between study sites (ManyPrimates et al. 2022), the classical DRT presents an unrealistic scenario in which monkeys generally perform poorly compared to their great ape relatives and commonly appear to forget hiding locations within less than a minute (Amici et al. 2010; ManyPrimates et al. 2022; Schubiger et al. 2016). Furthermore, recently it has been suggested that the classical DRT does not measure memory but rather taps into the domain of attention (van Schaik et al. 2024). However, slight adjustments to the test format can improve performance. For instance, common marmosets (*Callithrix jacchus*) and squirrel monkeys (*Saimiri sciureus*) performed poorly in a two-choice DRT without a clear trajectory of performing worse in conditions with a longer delay (Schubiger et al. 2016). Contrary to what one would expect, increasing task difficulty by decreasing the chance level of making a correct choice (nine cups instead of two) increased overall performance and revealed a typical forgetting curve with decreasing performance with increasing delays.

Another way to assess memory is in open foraging paradigms, simulating a more natural setting for the animals. These, "naturalistic" test scenarios provide evidence for impressive abilities to memorize food locations, even in small-brained species that usually perform poorly in the DRT (ManyPrimates et al. 2022; Menzel and Juno 1985).

In a series of experiments, saddle-back tamarins (*Saguinus fuscicollis*) could freely explore a testing enclosure that contained a set of baited and unbaited objects. The individuals could learn within one trial about the location and properties of an object that contained food and would revisit these objects after a delay of 24 h significantly earlier and faster than objects that previously did not contain food. In contrast to the classical DRT, this delayed memory task (24 h-DMT) showed that these small-brained tamarins could memorize food locations for an extended period after a single learning trial. Similar effects have been found in marmoset monkeys when assessing spatial memory in open-foraging tasks (Abreu et al. 2020; Easton et al. 2003; MacDonald et al. 1994; Platt et al. 1996; Vannuchi et al. 2020). In contrast to the DRT, such open foraging paradigms provide animals with both visual and spatial clues, which could improve performance. In fact, studies have shown that performances in tasks requiring spatial memory decline faster with age than performances in tasks relying on non-spatial memory (Barnes et al. 1987), which may be due to aging effects on the hippocampus (reviewed in Rosenzweig and Barnes 2003). Therefore, when relying solely on spatial memory, performance is expected to be lower compared to scenarios that utilize both spatial and visual memory. Furthermore, animals can explore cues for a longer time period than in the DRT, which may further consolidate memory. Hence, it is expected that animals perform better in open-foraging paradigms than in the classical DRT. However, it remains unknown whether individual performance in the 24 h-DMT correlates with that in the classical DRT, which would be expected if both tasks indeed measure memory.

Second, it is widely accepted that many cognitive abilities decline with aging (Salthouse 2019), and memory is no exception. In humans, memory performance constantly increases until it reaches its peak at around 20 years of age after which it declines linearly (Brockmole and Logie 2013). Similar age effects on memory have also been investigated quite extensively in animals (reviewed in Berchtold and Cotman 2009; Gallagher and Rapp 1997). One explanation for the decrease in cognitive performance is the so-called *neural dedifferentiation* in which neural processes become less selective and cognitive processes like memorizing information are impaired (Alexander et al. 2012; Berchtold and Cotman 2009; Koen et al. 2020). However, although *neural dedifferentiation* predicts this cognitive decline, findings from animal studies have yielded inconsistent results in this regard. For instance, whereas performance in pigeons in an associative learning task declined with age, memory was unaffected (Flaim and Blaisdell 2023). In a study with long-tailed macaques (*Macaca fascicularis*), age did predict memory performance in the delayed response task (Darusman et al. 2014). In contrast, in a recent comparison across 41 primate species, age improved memory performance

overall (ManyPrimates et al. 2022), which may be due to a lack of older individuals and a bias toward younger individuals with still increasing cognitive performance in this study. Alternatively, it may be that the conducted tests effectively did not examine memory, but instead measured a different cognitive component altogether that is not affected by increasing age.

Third, a component that will influence the result of memory tasks, or in fact any cognitive test, is attention toward the task at hand. Attention has been closely linked with cognitive performance in animals and humans (Decker and Duncan 2020; Schubiger et al. 2020). For instance, in great apes, it has been shown that establishing signals and securing the attention of the tested animals can increase cognitive performances (Mulcahy and Call 2009; Mulcahy and Sudendorff 2011). In another example, marmosets paid little attention when the costs of being unattentive were rather low, but increased their attentiveness when chances of being correct decreased (Schubiger et al. 2016). Especially in these smaller species where individuals inherently face a higher predation risk, individuals might pay more attention to their surroundings than to the actual test, which if not controlled for, can impair their cognitive performance (Schubiger et al. 2015).

In this study, we tested common marmosets in two different tasks examining memory performance. First, we tested individuals in an experimental design that resembles the 24-h DMT (Menzel and Juno 1985), followed by examining memory using the classical DRT. Marmosets are group-living, arboreal, and relatively small-brained platyrrhines (Solomon and Rosa 2014; Yamamoto et al. 2010) that show high levels of vigilance also in captivity (Brügger et al. 2023; Phaniraj et al. 2024). Their natural distribution ranges from the dry caatinga to the humid Atlantic forest (Garber et al. 2019). Marmosets are generalists feeding on a wide range of food items (Souto et al. 2007) for which they use a variety of foraging techniques. Generalists are thought to exhibit a superior spatial memory compared to specialists (Henke-von der Malsburg et al. 2020), and indeed wild common marmosets optimized their foraging route in a spatial memory task based on single-trial learning (Abreu et al. 2020; Xavier et al. 2024). Performance in spatial memory was observed over short-term (4.5 h) and long-term (11 days) in which marmosets approached previous food locations more often compared to locations that did not contain food. Hence, like the tamarins in the 24-h DMT (Menzel and Juno 1985), memory performance in a more natural setup surpassed the usual outcome of the classical DRT. However, a second crucial difference to the classical DRT is that in both studies (Abreu et al. 2020; Menzel and Juno 1985) individuals were tested within their social unit rather than alone. Not only can isolation manifest in increased stress (Ward and Webster 2016) impairing cognitive performance, but being in a group


can have positive effects on cognitive performances of problem-solving (Sehner et al. 2022), which are not restricted to, but may well include, finding food locations. Hence, the question remains whether the testing performance was better than expected since animals were not tested individually or because the experiment itself resembled a more natural setup.

This study aimed at examining the ecological and content validity, as well as the commensurability, of two commonly used tests. Therefore, we employed the 24 h-DMT and the classical DRT in common marmosets. We predicted that animals would excel in the more realistic setup of the 24 h-DMT, whereas performance in the classical DRT would be moderate at best and rapidly decrease with increasing latency. We furthermore explored which factors contributed to individual performance in both tasks. Following the general assumption that memory performance decreases with age, we expected that older individuals would do worse in both experiments if they indeed measured memory (Darusman et al. 2014). In addition, we explored whether sex could influence performance. Since previous studies have pointed out that males react stronger towards experimenters than females and are less attentive (Schubiger et al. 2015), we expected a better performance of females in the DRT but not in the 24 h-DMT, given that the latter was conducted without an experimenter present. Lastly, assuming that both tests examine memory we expected an intraindividual correlation between test performance scores.

## Methods

### Subjects and housing

In this study, we tested 16 (9 females) adult common marmosets (*Callithrix jacchus*) from five captive family groups. The age range of the individuals was from 1.5 to 18 years (for details see Table S1 in Online Resource 1). Older animals have experiences with cognitive tasks, whereas it was the first cognitive experiment for the three individuals under three years of age. However, the only individuals who have experience with memory tasks (the DRT) are Kapi and Vito. All animals were born in captivity and housed in heated indoor enclosures with ad libitum access to outdoor enclosures (if weather conditions allowed it). Both, indoor and outdoor enclosures were enriched with natural groundcovers and natural climbing structures. In addition, all enclosures included additional hiding opportunities as well as access to heating lamps. Animals were fed several times a day with a vitamin-enriched porridge in the morning, a mixture of fruits and boiled vegetables over midday, and a snack composed of either animal protein (e.g. fish or boiled egg) or

 Springer



whether the animal had found the food. The order in which the animals within a family were tested was random and depended on the motivation of the individuals. Animals could enter a system of tubes attached to their home enclosure and walk to the test enclosure. The tube system ended in the small cage inside the test enclosure. After an animal entered the small cage the tube system was blocked for the duration of the experiment. To enter the test enclosure, the experimenter lifted a small guillotine door, which was lowered once the animal entered the test enclosure. After the session, the experimenter lifted the guillotine door again and the animal could return to its respective home enclosure using the tube system. Before the next animal was tested, we cleaned food leftovers and rebaited the object if necessary.

### Data scoring

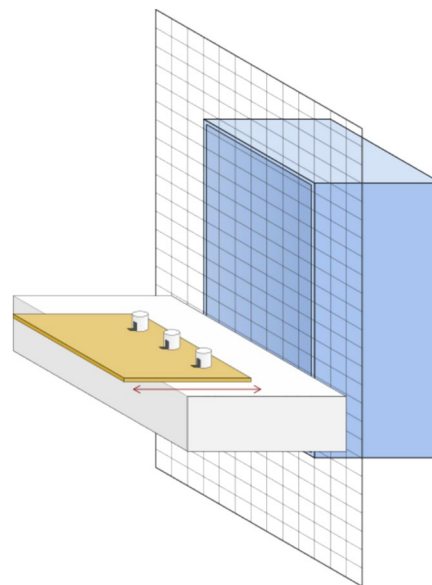
All trials were video recorded with two GoPro Hero 7 White and analyzed with INTERACT software V18.1.4.4 from MANGOLD GmbH. The two cameras were positioned on opposite sides of the enclosure (Fig. 1), to monitor the whole test enclosure. Each trial started on the preceding frame when the animal started to move away from the guillotine door and entered the test enclosure, and lasted for 15 min. We coded the order in which the animal interacted with each object. We also noted the latency for the first interaction with a currently-baited object. An interaction was defined as touching a test object or sniffing/watching the object from a distance up to one arm-length. The coding system consisted of two levels. First, we coded which object was interacted with and then we coded the baiting state of that object of that particular session. A choice was counted as correct choice if an animal interacted first with the baited object before it approached any other object. A second observer coded 20% of all second trials determining when and with which object an individual interacted first. Inter-observer reliability was assessed using Cohen's Kappa, which yielded a value of 0.83, indicating almost perfect agreement between the two observers. This result was statistically significant ( $z = 5.85$ ,  $p < 0.001$ ).

### Experiment 2: Classical delayed response task (DRT)

After the 24 h-DMT, we tested the same 16 individuals using the classical DRT (see also ManyPrimates et al. 2022). Animals were tested individually in a compound (80 cm height  $\times$  40 cm width  $\times$  40 cm depth), inside their home enclosure. The compound (henceforth "test compound") was attached to the inside of the home enclosure front grid. During periods between experimental sessions, animals had ad libitum access to the test compound. Thus, animals could habituate to the test compound during the entire time unless another individual was tested. To avoid interference,

the experimenter moved the group members who were not tested through a tube system to an adjacent room where they had food, water, and enrichment items.

The general setup included a white table suspended by two metal hooks in front of the test compound and outside the home enclosure (Fig. 2). The white table was affixed to the enclosure only during test trials and was removed once a session was completed. Attached to the top of the table was a wooden board (40 cm width  $\times$  40 cm depth), and three white opaque cylinder-shaped cups (5 cm high  $\times$  3 cm in diameter) were placed on top of the board. The cups were fixed with the opening downwards to maintain the same even distance throughout the test. The board was attached to a pair of drawer tracks and could be moved out of the subject's reach. The experimenter stood behind the board and turned the three cups with the opening facing the subject. Next, the experimenter showed the bait to the subject and placed it in front of one of the cups. The position could be (1) left, (2) middle, or (3) right – from the experimenter's perspective. After that, the experimenter erects the cups from left to right, thereby hiding the food item. The delay conditions were short (0-s delay), medium (15-s delay), or long (30-s delay). After the respective delay, the experimenter slid the board towards the subjects while avoiding giving any clues about the hiding position. The marmoset could then choose by pointing toward or touching one of the cups through the grid. If the individual made an ambiguous choice, the trial



**Fig. 2** Schematic drawing of the test compound. The front grid represents the enclosure grid, and the test compound inside the enclosure was made of a grid, and the blue pattern differentiates it from the enclosure's grid. The wooden table can be moved back and forth. The cups were held with a small stripe of tape so as not to fall and interfere with the experiment. See the online article for the color version of this figure

was not counted and repeated. An ambiguous choice was considered when an animal reached in between two cups and/or did not touch any cup. The trial was also repeated if the subject did not choose within one minute. After the marmoset chose a cup, the board was pulled back, and the chosen cup was lifted and laid down. If the cup revealed the bait, the subject received it as a reward. If the cup did not contain the bait, the subject did not receive a reward, and the same food item was used in the subsequent trial. After revealing the chosen cup, the experimenter lifted the other cups, and a new trial was started.

## Training

Before starting the test sessions, each subject was trained to ensure reliable choice behaviors. This procedure was not required by the ManyPrimates et al. (2022) project but advised, as the animals had to consistently point to, or reach for, visible food items placed on the table board before the start of the experiments. Accordingly, subjects went through a pre-test phase in which the experimenter followed the same procedure but did not impose any time delay. After the subject reached the criterion ( $\geq 70\%$  correct choices within 12 consecutive trials) it entered the test phase on the next day. The duration of training varied between subjects, but all 16 individuals passed the criterion of making the correct choice in 70% or more in 12 consecutive trials without delay. Correct choices were those where the animal chose the baited container. In contrast to the ManyPrimates et al. (2022) project, the subject received one warm-up trial without delay at the beginning of each test day. This trial was not considered and served only to attract the subject's attention as well as to probe its motivation to participate on a given day. Two predefined criteria for stopping a session were used: (1) the subject did not make a choice in one minute in three consecutive trials, or (2) the subject was no longer attentive to the table and/or emotionally aroused and indicated it wanted to leave the test compartment. If the subject showed at least one of these criteria, we stopped the session immediately and restarted the trial the following day.

The ManyPrimates et al. 2022 project provided the hiding location and delay conditions (see Table S2 in Online Resource 1). The hiding position was randomized across trials with the following constrictions: the exact location could not occur more than two consecutive times, and each hiding location occurred an equal number of times per delay. Trials were grouped in blocks, with each block comprising three trials of the same delay (short, medium, and long). The number of trials per session varied among individuals, with the constraint that there were at least three trials per test day. If less than three trials were finished, the session was terminated, and trials were discarded and repeated the following day.

## Data scoring

All trials were video recorded with a GoPro Hero 9 Black, scored live using ManyPrimates et al. (2022) check sheets, and confirmed later with the video. We coded with which cup a subject interacted first after the board slid forward. We also coded the position of the indicated cup (left, middle, or right). If the indicated cup hid the food, the response was coded as correct. A second person scored 20% of all trials scoring the choice of an animal. The inter-observer reliability was assessed using Cohen's Kappa, which yielded a Kappa value of 0.93, indicating almost perfect agreement between the two observers. This result was statistically significant ( $z = 13.7$ ,  $p < 0.001$ ).

## Statistical analysis

All statistical analyses were performed in R-studio (R Core Team 2020). To test whether animals remember the location of the last baited object in the 24 h-DMT, we ran a binomial Generalized Linear Mixed-effects Model (GLMM) with a log link-function. The dependent variable scored whether the first interaction was with the currently baited object or not (yes/no). The model did not include fixed effects, however, *individual identity* nested in *family identity*, and *session number* as random effects. We also included the probability of making a correct choice by chance as an offset term. The offset was defined as the log of 1 over  $n$ , where  $n$  was the total number of objects in each session, and thus varied from 1/3 (first session) to 1/30 (last session). In addition, we conducted a Monte Carlo simulation generating 10,000 random choices for each session amounting to 100,000 random choices for all sessions. Again, we adjusted the probability of making a correct choice by chance level from 1/3 in the first session, to 1/30 in the last session. We furthermore matched the number of animals in the simulation to the empirically observed number of animals that found the baited object in our experiment in the first session. For instance, whereas for the first session, we simulated 10,000 random choices by 16 animals, the last session was based on 10,000 random choices by 11 animals. Thus, the subsequent comparison of the simulated data to the sum of observed first interactions with baited objects on second trials is much more objective.

We formulated a Cox proportional hazards mixed-effects model for the *latency of the first interaction* with the baited object to examine whether animals would find the baited object faster in their respective second trials. The model included the *individual identity* nested in *family identity* and *session number* as random factors, and *trial* as a predictor variable.

To examine whether individual characteristics affected performance, we calculated a binomial GLMM including *age*, *sex*, and *status* as predictor variables, and *session*

*number* and *individual identity nested in family identity* as random effects. Our response variable was structured as a two-column matrix: one column indicated the *number of correct choices* and the other indicated the *number of failures*.

To test whether animals would remember the location of the hidden food item in the classical DRT, we ran a binomial GLMM with a log-link function. First, we examined how individuals performed over all trials, regardless of the delay. The model did not include fixed effects but the family identity as a random effect. We also included the probability of making a correct choice by chance (1/3) as an offset term. In a subsequent analysis, we tested whether the *delay condition* affected the performance. For this, we set the *delay* as a contrast and included it as a predictor in the model. Lastly, we tested whether animals were better than expected by chance for each *delay condition*, calculating a binomial GLMM with the same structure as used for the overall performance.

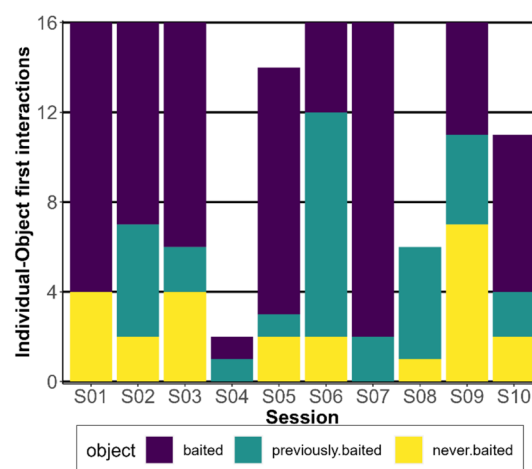
Similar to the first experiment, we examined whether individual characteristics affected performance by calculating a binomial GLMM including *age*, *sex*, and *status* as predictor variables and *family identity* as random effect. Our response variable was structured as a two-column matrix: one column indicated the *number of correct choices* and the other indicated the *number of failures*.

## Results

Overall, we recorded 320 trials in the 24 h-DMT, equally divided into 160 trials of first and second trials over all sessions. In 80% of all the trials the individual found the baited object, resulting in 256 successful trials (see Table S1 in Online resource 1).

We examined whether animals would remember the location after one trial learning and switch from previously baited objects in previous sessions to novel objects that are baited (Fig. 3). We performed an analysis restricting correct choices to currently baited objects only. The intercept was significantly positive (GLMM: estimate = 2.55; standard error = 0.47,  $z$  value = 5.46;  $p < 0.001$ ), indicating that individuals interacted more frequently first with the currently baited object (73 out of 129 trials) than would be expected by chance alone. We also calculated the number of correct choices in each session to control for difficulty, since the increasing number of objects may have increased the difficulty level in each following session. Our results demonstrated no increase or decrease in the number of correct choices throughout the experiment (see Table S3 in the Online Resource 1 for the table of polynomial contrasts).

We further examined whether the observed behavioral pattern was likely the result of individual memory and one-trial learning or random choice. We compared the random



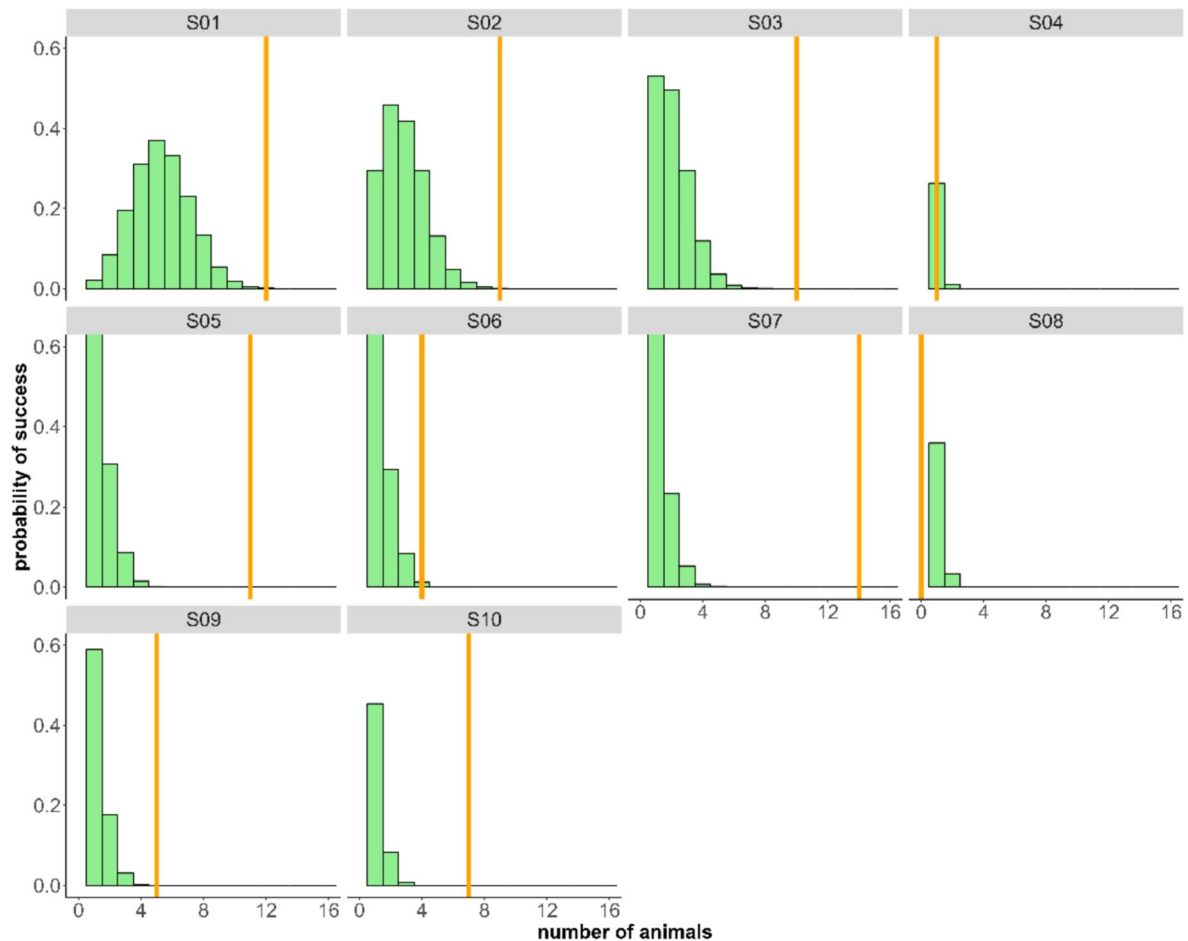
**Fig. 3** The number of individuals who found the food in the second trial of each session, separated by their first interaction with a baited object (purple), a previously baited object (turquoise), or a never baited object (yellow) in their respective second trial. See the online article for the color version of this figure

choices extracted from the Monte-Carlo simulation with the observed first choices of a baited object (Fig. 4). This analysis thus did not consider previously baited objects. In every session, except for sessions four and eight, the observed correct choices were significantly higher than the simulated ones (Table 1). On average, the observed value is  $2.15 \pm 1.42$  times bigger than the 97.5% theoretical quantile.

In the next step, we examined whether animals would find the correct object faster in the second trial than in the first trial of a session. Hence, we compared the latencies of finding the correct object in the first trials with the latencies of finding the correct object in the respective second trials. A Cox mixed effect model revealed that an animal found the baited object earlier during its second trial than on its first trial (proportional hazard ratio = 1.91,  $CI_{95\%} = 1.46$  to 2.49,  $p < 0.001$ ). In other words, at any given instant in time, an individual was 1.91 times more likely to already have found the correct object in the second trial than in the first trial (Fig. 5).

Lastly, we examined whether individual differences like age, sex and status (i.e. being a helper vs a breeder) were associated with performance. Our model including only first interactions with the baited object as a correct choice revealed no effect of *age*, *sex* and *status* and was not significantly better than the null model ( $X^2(1) = 5.09$ ,  $p = 0.16$ ; see Table S4 in the Online Resource 1 for the full model summary).

For the DRT, we conducted a total of 543 trials. 14 subjects completed all 36 trials each, one female completed 27 trials, and one male completed 12 trials. The binomial GLMM revealed a significant positive intercept (GLMM: estimate = 0.57; standard error = 0.22,  $z$  value = 2.5:



**Fig. 4** Simulated probability of the number of animals interacting with the correct object (currently baited object) based on random interactions (green histogram). The simulated data for each session is based on the number of animals that found the correct object in the first trial. The orange bar represents the observed number of animals

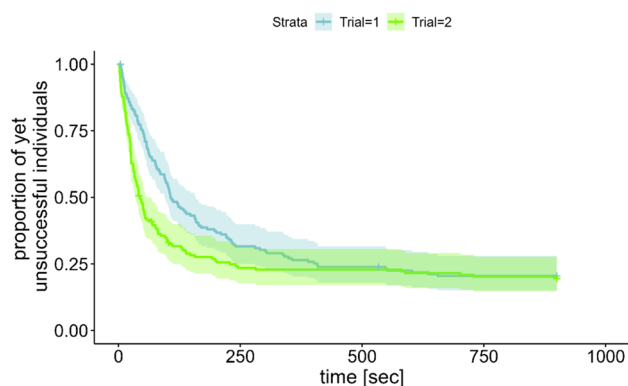
interacting first with the correct object in the second trial. In eight out of ten sessions, a significantly higher number of animals than expected by chance interacted first with the currently baited object (see also Table 1). See the online article for the color version of this figure

**Table 1** 95% quantiles and the median of the simulated distribution of the number of animals that would interact correctly by chance per session

Session	Number of considered trials	2.50%	50%	97.25%	N	p-value
S01	16	2	5	9	12	<b>&lt; 0.001</b>
S02	16	0	3	6	9	<b>&lt; 0.001</b>
S03	16	0	2	4	10	<b>&lt; 0.001</b>
S04	2	0	0	1	1	0.15
S05	14	0	1	3	11	<b>&lt; 0.001</b>
S06	16	0	1	3	4	<b>&lt; 0.01</b>
S07	16	0	1	3	14	<b>&lt; 0.001</b>
S08	6	0	0	1	0	1
S09	16	0	0	2	5	<b>&lt; 0.001</b>
S10	11	0	0	2	7	<b>&lt; 0.001</b>

Observed number of animals that first interacted with the baited object (N) and the p value compared to the simulated distribution. Bold values represent  $< 0.05$ . Considered are those trials 2 for each session where the individual found the bait during trial 1

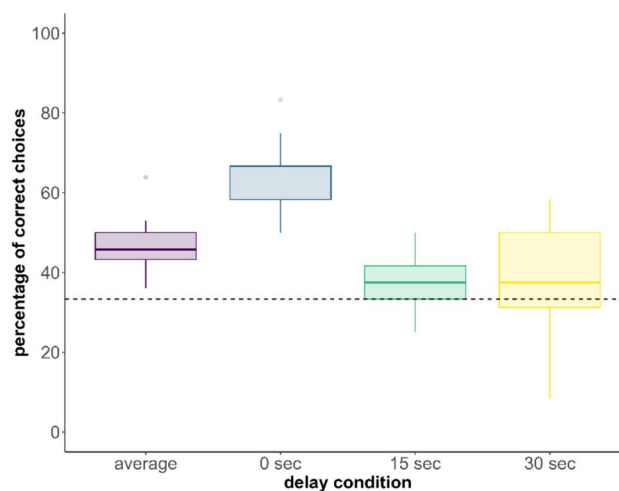




**Fig. 5** Kaplan–Meier plot comparing the latencies of finding the correct object in both trials of a session. Individuals found the correct object significantly faster during their second trial (green) than their first trial within the same session. See the online article for the color version of this figure

$p=0.01$ ), suggesting that across all conditions individuals performed significantly above chance level. Subsequently, we examined whether there was a difference in performance between the different delay conditions. The full model was better than the null model ( $X^2(1)=34.96$ ,  $p<0.001$ ) and revealed a significant linear decrease in performance with increasing delay (GLMM: estimate =  $-0.81$ ; standard error =  $0.15$ ,  $z$  value =  $-5.2$ ;  $p<0.001$ ). Furthermore, the quadratic contrast of the delay condition also significantly influenced the performance (GLMM: estimate =  $0.39$ ; standard error =  $0.15$ ,  $z$  value =  $2.52$ ;  $p=0.01$ ). Altogether the results suggest that the decline in performance is not constant with increasing delay but rather indicates a step decrease from the 0 s delay condition to the 15 s delay condition and almost no difference between the 15 s delay and the 30 s delay (Fig. 6). Furthermore, a subsequent analysis revealed that individuals did not perform above chance in either the 15-s or the 30-s delay scenario. Only when the delay was 0 s did animals perform better than expected by chance (GLMM: estimate =  $1.3$ ; standard error =  $0.22$ ,  $z$  value =  $2.5$ ;  $p=0.01$ ).

An important factor to consider for such tests is whether animals developed a side bias and preferably chose one cup more often than another regardless of the manipulation by the experimenter. We defined that animals have a side bias if they choose one cup more often than in one-third of all trials plus one standard deviation (in this case  $12+1SD$ ). The standard deviation was calculated by simulating random choices of 16 animals over 36 trials (12 and 27 trials for the individuals that did not complete all sessions respectively). Among the 16 tested animals, we identified 11 with a side bias (see Fig. S2 in the Online Resource 1). The marmosets that were considered biased chose the correct cup 46.1% ( $SD=7.75\%$ ) of the time, while the non-biased chose the correct cup 47% ( $SD=3\%$ ). Individuals with a side bias did not significantly



**Fig. 6** Boxplot of the percentages of correct choices separated into delay condition and average of all conditions. On average animals performed better than expected at chance-level (dotted line). However, this pattern was driven by the shortest delay as in the 0-s delay condition but not in the 15-s or 30-s condition, animals performed significantly better than by chance

perform worse than individuals without a side bias (GLMM: estimate =  $-0.04$ ; standard error =  $0.18$ ,  $z$  value =  $-0.21$ ;  $p=0.83$ ).

The model containing *delay*, *age*, *sex*, and *status* as predictor variables and *family identity* as a random variable was significantly better than the null model ( $X^2(1)=11.13$ ,  $p=0.02$ ; see Table S5 in the Online Resource 1 for the full model summary). However, only delay had a significant effect on the performance, and neither age, sex nor status increased or decreased the model fit.

### Comparing the results of the 24 h-DMT and the DRT

Ultimately, our objective was to examine a potential correlation between the number of correct choices observed in the 24 h-DMT experiment and the DRT experiment (see Fig. S3 in the Online Resource 1). Using a Spearman's correlation test, we found no correlation between the performances of the 24 h-DMT experiment with the short delay DRT ( $\rho=-0.41$ ;  $z=-1.61$ ;  $p=0.11$ ), the middle delay ( $\rho=-0.21$ ;  $z=0.79$ ;  $p=0.43$ ) or the long delay ( $\rho=-0.22$ ;  $z=0.84$ ;  $p=0.4$ ). Furthermore, there was no correlation between the 24 h-DMT experiment and the average performance of the DRT experiment ( $\rho=-0.12$ ;  $z=-0.47$ ;  $p=0.64$ ).

### Discussion

We examined memory capabilities of 16 marmoset monkeys using an open-foraging paradigm, the 24 h-DMT, and the classical DRT and compared both results with each other.

For the 24 h-DMT, subjects were released individually into an unfamiliar environment containing up to 30 objects, with only one object containing food. Following a 24-h delay, we assessed whether the monkeys retained the memory of the food's location. In the second experiment we used the approach of the classical DRT. In the following we will first discuss the results of the 24 h-DMT, followed by the results of the classical DRT and the comparison of both. Lastly, we will discuss why results obtained from the classical DRT should be interpreted cautiously, as experiments conducted with more naturalistic setups can reveal different memory capacities, offering an alternative perspective on animals' cognitive abilities.

Our findings from the 24 h-DMT align with previous research assessing spatial memory in marmoset monkeys. Marmosets perform well in open foraging tasks in both captivity (Easton et al. 2003; MacDonald et al. 1994; Platt et al. 1996; Vannuchi et al. 2020) and the wild (Abreu et al. 2020) and use heuristic rules to move between food locations (Xavier et al. 2024). In addition, like tamarins (Menzel and Juno 1985), marmosets were able to remember food locations for 24 h after a single trial of exploration. Although marmosets are omnivorous generalists, feeding on mobile prey like insects and small vertebrates, they also rely on (spatially and temporally) stable food sources like tree exudates (Souto et al. 2007). A substantial part of their diet comprises gum, and their home ranges are typically associated with the availability of gum trees. Their movement patterns often involve traveling between sleeping trees and gum trees as part of their daily activities (reviewed in Schiel and Souto 2017). Hence, remembering the location of stable food sources like gum trees is important in the ecology of common marmosets. Our results also confirm that marmosets tend to rely on a *win-stay* strategy (MacDonald et al. 1994), meaning that they prefer to forage at locations where they have remembered being successful in the past. Food locations did not change between the first and the second trial of a session and marmosets tend to revisit the location where they found food during the last trial rather than during previous sessions. Their tendency to explore previously baited objects over novel objects first suggests a strong inclination toward the *win-stay* strategy. We also confirmed that marmosets would visit the baited object faster in their second trial of a session, corroborating that they remembered the location from the first session.

The results from the classical DRT follow the general trend of studies using this paradigm showing that an increase in the delay reduces the proportion of correct choices (Amici et al. 2010; Barth and Call 2006; Call 2001; Darusman et al. 2014; ManyPrimates et al. 2022; Schubiger et al. 2016). The marmosets performed rather poorly and performed at chance level at a delay of 15 s and 30 s. However, even in the no-delay condition which requires no memory, the performance

of marmosets was below the average performance of other primate species in the same task (ManyPrimates et al. 2022). Given that the individuals were trained prior to the experiment, it is improbable that their performance is due to a lack of understanding of the task. Although we did not assess attention, it remains a likely explanation that the animals did not pay much attention to the experimenter and the task at hand and rather chose a side bias strategy which was on average successful in 33% of all trials. In fact, individuals with a side bias achieved the same number of rewards as those without. It has been shown that marmosets (and squirrel monkeys) pay more attention to such tasks when the probability of being successful is smaller (Schubiger et al. 2016). Similar results of such poor performances have been found in previous studies using marmosets as role models (Miles 1957). In this classical study by Miles (1957), marmosets required over 1200 trials to perform at the same level as macaques in a 16-s delay DRT. Among other potential explanations, such as fundamental differences in their brain structure, it was speculated that marmosets exhibited poor attention to the task. In a similar study where the cues were given via a touch screen and after close to a hundred or more trials, four out of six marmosets made a correct decision after a 100-s delay (Yamazaki et al. 2016). Similar to our study, the marmoset tested was physically separated from other individuals but still at least in hearing distance with at least one conspecific. Two key differences between our study and that of Yamazaki et al. (2016) may explain the variation in performance. First, the individuals tested in our study had less overall training and fewer experimental trials. Previous research suggests that marmosets improve their performance over hundreds of trials, eventually reaching nearly 90% accuracy (Miles 1957; Yamazaki et al. 2016). An alternative explanation for improved memory could be that marmosets become well habituated to the test scenario, leading to reduced vigilance after hundreds of trials. This reduced vigilance may allow them to focus more attention on the stimuli. Second, in our study, clues were provided by an experimenter rather than automatically. Even minor variations in an experimenter's behavior, despite standardized protocols, can influence an animal's response.

Next, we examined the effect of individual traits on the overall performance of both experiments. Neither in the 24 h-DMT, nor in the classical DRT could we pick up any factors that would explain individual differences. Against our prediction that performance would decline with age, we found no significant effect of age. Results on the effect of age on memory in non-human animals are mixed and sometimes reveal a significant decline (Darusman et al. 2014) or no effect (Flaim & Blaisdell 2023; ManyPrimates et al. 2022). One possible explanation for this inconsistency is the influence of testing experience. Older individuals may compensate for their cognitive decline with testing experience.

Although the here tested individuals have not been tested in either the 24 h-DMT or the classical DRT (with the exception of Kapi and Vito (Schubiger et al. 2016); see Table S1 in Online Resource 1) before, experience from other cognitive tests may still facilitate their performance. Older individuals have participated in previous experiments (Brügger et al. 2021; Miss and Burkart 2018; Schubiger et al. 2015; Sehner et al. 2022), whereas it was the first testing experience for younger animals. We also expected a better performance of females over males based on previous results from this colony (Schubiger et al. 2015). However, sex had no effect on the overall performance in both experiments, which may be due to the same reasoning as for age. The females in our dataset had less testing experience than the males. In addition, males are more aroused than females during tests in the presence of an experimenter (Schubiger et al. 2015), which in our case could only have had an effect on the second but not the first experiment.

Finally, we examined whether the performances of the two experiments correlate with each other. Overall, the performance in the DRT was much worse than in the 24 h-DMT and we found no correlation between the tasks. Previous experiments have shown that information can be temporarily inaccessible and that memory performances after 24 h and a sleeping period in between are better compared to shorter delays (Martin-Ordas and Call 2011). This may be a contributing factor to the good performance during the 24 h-DMT compared to the DRT but we argue that it cannot explain why the average performance in the 24 h-DMT (80% correct choices) is better than the average performance in the no-delay condition of the DRT (65% correct choices). A second difference between the two experiments was the presence of an experimenter. Whereas the DRT required the presence of an experimenter at all times, the 24 h-DMT was mostly performed in the absence of a person. Especially smaller species like marmosets can be influenced by the presence of an experimenter and attention to the task is impaired (Schubiger et al. 2015). Hence, marmosets in the DRT are likely to be more distracted than individuals in the 24 h-DMT potentially decreasing performance. Lastly, the available information for the individuals was higher in the 24 h-DMT compared to the DRT. In the former experiment, marmosets had access to spatial and visual information and could explore the locations of the hidden food reward longer than in the DRT. Again, spatial memory seems to decline more rapidly than other forms of memory (Rosenzweig and Barnes 2003), meaning that performance is expected to be lower in the DRT compared to the 24 h-DMT.

These results have implications for studies examining memory, in particular studies that aim at comparing between multiple species (e.g., Amici et al. 2010; ManyPrimates et al. 2022). The classical DRT represents one of the most repeated tests to assess memory, especially in

primates (Amici et al. 2010; Barth and Call 2006; Darusman et al. 2014; Herrmann et al. 2007; ManyPrimates et al. 2022; Schubiger et al. 2016). Despite its widespread application, our results raise doubts about its validity and call for adjustment of this paradigm, at least for marmosets and more likely for other species as well. Previous studies have shown that animals perform better when the experimenter secures the attention of the tested individual. In great apes, attention can be secured via both eye contact (Mulcahy and Call 2009; Mulcahy and Suddendorf 2011) and ostensive signals (Kano et al. 2018). However, depending on the social structure, some species may react threatened to eye contact or do not respond to ostensive signals (Harrod et al. 2020). Another possibility to achieve increased attentiveness and thus more accurate measurement of memory performance is to decrease the probability of success in such tests (Schubiger et al. 2016). Common marmosets and squirrel monkeys paid more attention to the test when the potential costs of not paying attention were higher. However, the overall performance of non-human primates in these classical DRTs remains rather poor or moderate at best, showing that even great apes have problems locating the right container after only a couple of seconds (ManyPrimates et al. 2022).

An alternative and less artificial way of assessing memory is to use open foraging paradigms. Compared to the classical DRT, performances in these open-foraging paradigms are less affected by delay and correct choices remain above chance level in conditions of shorter (Platt et al. 1996) and longer delay (Abreu et al. 2020; Menzel and Juno 1985; but see Platt et al. 1996). Additionally, open foraging paradigms can be conducted in a way that the presence of an experimenter is unnecessary and without the necessity to train animals on an automatic apparatus, thus avoiding potential interferences of different individuals toward one or multiple experimenters (Schubiger et al. 2015). Open-foraging paradigms also seem to circumvent the problem of motivational biases mentioned by Schubiger et al. (2016) since animals perform well above chance with only a few items to choose from (Abreu et al. 2020) and multiple items as seen here and in previous results (Menzel and Juno 1985). Although the chance level cumulatively decreased with increasing sessions and was ultimately much lower than in the classical DRT, performance in the 24 h-DMT did not decrease. In addition, individuals can collect and memorize information rather independently as they would under natural conditions.

This undermines the capacity to memorize food locations in common marmosets if tested in a paradigm that has some ecological relevance. Marmoset diet heavily relies on exudates (Schiel and Souto 2017) and marmosets have to remember where these food sources are located. Hence, like in the 24 h-DMT paradigm, they freely explore their environment and memorize spatial and visual information about food locations. Finally, unlike the classical DRT,

open-foraging paradigms can be applied to wild populations (Abreu et al. 2020), thus allowing us to test memory under natural conditions.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10071-025-01944-3>.

**Acknowledgements** We are grateful to Gisepp Bazzell for taking care of the animals and assisting with their handling. Thanks to Eloisa Martins for her support in creating the images of the apparatuses and test enclosure. We are grateful to Rodolph Schlaepfer for his help with the camera setup. This project was supported by an SNF Grant (grant number 31003A\_172979) to Judith Burkart and an SNF Postdoc. Mobility Grant (grant number P500PB\_217864/1) to Sandro Sehner.

**Author Contribution** S.S., F.M., and J. M. B. designed the study. F.M. collected the data. S.S., F.M., and E. P. W. analyzed the data. S.S., F.M., and J. M. B. wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This project was supported by an SNF Grant (grant number 31003A\_172979) to Judith Burkart and an SNF Postdoc. Mobility Grant (grant number P500PB\_217864/1) to Sandro Sehner.

**Data availability** All data needed to evaluate the conclusions in the paper are present on OSF (<https://doi.org/10.17605/OSF.IO/3V6SZ>).

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Ethical approval** All the experiments were in accordance with the Swiss legislation and licensed by the Kantonales Veterinäramt Zürich (license number: ZH232/19, degree of severity: 0).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abreu F, Souto A, Schiel N (2020) Wild common marmosets (*Callithrix jacchus*) employ spatial cognitive abilities to improve their food search and consumption: an experimental approach in small-scale space. *Primates* 61(6):807–816. <https://doi.org/10.1007/s10329-020-00826-1>
- Alexander GE, Ryan L, Bowers D, Foster TC, Bizon JL, Geldmacher DS, Glisky EL (2012) Characterizing cognitive aging in humans with links to animal models. *Front Aging Neurosci*. <https://doi.org/10.3389/fnagi.2012.00021>
- Amici F, Aureli F, Call J (2010) Monkeys and apes: are their cognitive skills really so different? *Am J Phys Anthropol* 143(2):188–197. <https://doi.org/10.1002/ajpa.21305>
- Banerjee K, Chabris CF, Johnson VE, Lee JJ, Tsao F, Hauser MD (2009) General intelligence in another primate: individual differences across cognitive task performance in a new world monkey (*Saguinus oedipus*). *PLoS ONE* 4(6):e5883. <https://doi.org/10.1371/journal.pone.0005883>
- Barnes CA, Green EJ, Baldwin J, Johnson WE (1987) Behavioural and neurophysiological examples of functional sparing in senescent rat. *Can J Psychol* 41:131
- Barth J, Call J (2006) Tracking the displacement of objects: a series of tasks with great apes (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Pongo pygmaeus*) and young children (*Homo sapiens*). *J Exp Psychol Anim Behav Process* 32(3):239–252. <https://doi.org/10.1037/0097-7403.32.3.239>
- Berchtold NC, Cotman CW (2009) Normal and pathological aging: from animals to humans. In: Bizon JL, Woods A (eds) *Animal models of human cognitive aging*. Humana Press, pp 1–28. [https://doi.org/10.1007/978-1-59745-422-3\\_1](https://doi.org/10.1007/978-1-59745-422-3_1)
- Brockmole JR, Logie RH (2013) Age-related change in visual working memory: a study of 55,753 participants aged 8–75. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2013.00012>
- Brügger RK, Willems EP, Burkart JM (2021) Do marmosets understand others' conversations? A thermography approach. *Sci Adv* 7(6):eabc8790. <https://doi.org/10.1126/sciadv.abc8790>
- Brügger RK, Willems EP, Burkart JM (2023) Looking out for each other: coordination and turn taking in common marmoset vigilance. *Anim Behav* 196:183–199. <https://doi.org/10.1016/j.anbehav.2022.11.007>
- Call J (2001) Object permanence in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and children (*Homo sapiens*). *J Comp Psychol* 115(2):159–171. <https://doi.org/10.1037/0735-7036.115.2.159>
- Darusman HS, Call J, Sajuthi D, Schapiro SJ, Gjedde A, Kalliokoski O, Hau J (2014) Delayed response task performance as a function of age in cynomolgus monkeys (*Macaca fascicularis*). *Primates* 55(2):259–267. <https://doi.org/10.1007/s10329-013-0397-8>
- Decker AL, Duncan K (2020) Acetylcholine and the complex interdependence of memory and attention. *Curr Opin Behav Sci* 32:21–28. <https://doi.org/10.1016/j.cobeha.2020.01.013>
- Easton A, Parker K, Derrington AM, Parker A (2003) Behaviour of marmoset monkeys in a T-maze: comparison with rats and macaque monkeys on a spatial delayed non-match to sample task. *Exp Brain Res* 150(1):114–116. <https://doi.org/10.1007/s00221-003-1409-5>
- Flaim M, Blaisdell AP (2023) The effect of age on delay performance and associative learning tasks in pigeons. *Learn Behav* 51(3):281–294. <https://doi.org/10.3758/s13420-022-00565-x>
- Gallagher M, Rapp PR (1997) The use of animal models to study the effects of aging on cognition. *Annu Rev Psychol* 48(1):339–370. <https://doi.org/10.1146/annurev.psych.48.1.339>
- Garber PA, Caselli CB, McKenney AC, Abreu F, De la Fuente MF, Araújo A, de Fatima Arruda M, Souto A, Schiel N, Bicca-Marques JC (2019) Trait variation and trait stability in common marmosets (*Callithrix jacchus*) inhabiting ecologically distinct habitats in northeastern Brazil. *Am J Primatol* 81(7):e23018. <https://doi.org/10.1002/ajp.23018>
- Harrod EG, Coe CL, Niedenthal PM (2020) Social structure predicts eye contact tolerance in nonhuman primates: evidence from a crowd-sourcing approach. *Sci Rep* 10(1):6971. <https://doi.org/10.1038/s41598-020-63884-x>
- Henke-von der Malsburg J, Kappeler PM, Fichtel C (2020) Linking ecology and cognition: does ecological specialisation predict cognitive test performance? *Behav Ecol Sociobiol* 74(12):154. <https://doi.org/10.1007/s00265-020-02923-z>



- Herrmann E, Call J, Hernández-Lloreda MV, Hare B, Tomasello M (2007) Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* 317(5843):1360–1366. <https://doi.org/10.1126/science.1146282>
- Hunter WS (1913) The delayed reaction in animals and children. *Anim Behav Monogr*. <https://doi.org/10.1037/11304-051>
- Kano F, Moore R, Krupenye C, Hirata S, Tomonaga M, Call J (2018) Human ostensive signals do not enhance gaze following in chimpanzees, but do enhance object-oriented attention. *Anim Cogn* 21(5):715–728. <https://doi.org/10.1007/s10071-018-1205-z>
- Koen JD, Sroková S, Rugg MD (2020) Age-related neural dedifferentiation and cognition. *Curr Opin Behav Sci* 32:7–14. <https://doi.org/10.1016/j.cobeha.2020.01.006>
- Lazaro-Perea C, Snowden CT, de Fátima Arruda M (1999) Scent-marking behavior in wild groups of common marmosets (*Callithrix jacchus*). *Behav Ecol Sociobiol* 46(5):313–324. <https://doi.org/10.1007/s002650050625>
- MacDonald SE, Pang JC, Gibeault S (1994) Marmoset (*Callithrix jacchus*) spatial memory in a foraging task: Win-stay versus win-shift strategies. *J Comp Psychol* 108(4):328–334. <https://doi.org/10.1037/0735-7036.108.4.328>
- ManyPrimates et al (2022) The Evolution of Primate Short-Term Memory. *Anim Behav Cogn*. <https://www.animalbehaviorandcognition.org/article.php?id=1334>
- Martin-Ordas G, Call J (2011) Memory processing in great apes: the effect of time and sleep. *Biol Lett* 7(6):829–832. <https://doi.org/10.1098/rsbl.2011.0437>
- Menzel E, Juno C (1985) Social foraging in marmoset monkeys and the question of intelligence. *Philos Trans R Soc Lond b, Biol Sci* 308(1135):145–158. <https://doi.org/10.1098/rstb.1985.0016>
- Miles RC (1957) Delayed-response learning in the marmoset and the macaque. *J Comp Physiol Psychol* 50(4):352–355. <https://doi.org/10.1037/h0040257>
- Miss FM, Burkart JM (2018) Corepresentation during joint action in marmoset monkeys (*Callithrix jacchus*). *Psychol Sci* 29(6):984–995. <https://doi.org/10.1177/0956797618772046>
- Mulcahy NJ, Call J (2009) The performance of bonobos (*Pan paniscus*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*) in two versions of an object-choice task. *J Comp Psychol* 123(3):304–309. <https://doi.org/10.1037/a0016222>
- Mulcahy NJ, Suddendorf T (2011) An obedient orangutan (*Pongo abelii*) performs perfectly in peripheral object-choice tasks but fails the standard centrally presented versions. *J Comp Psychol* 125(1):112–115. <https://doi.org/10.1037/a0020905>
- Phaniraj N, Brügger RK, Burkart JM (2024) Marmosets mutually compensate for differences in rhythms when coordinating vigilance. *PLoS Comput Biol* 20(5):e1012104. <https://doi.org/10.1371/journal.pcbi.1012104>
- Platt ML, Brannon EM, Briesse TL, French JA (1996) Differences in feeding ecology predict differences in performance between golden lion tamarins (*Leontopithecus rosalia*) and Wied's marmosets (*Callithrix kuhli*) on spatial and visual memory tasks. *Anim Learn Behav* 24(4):384–393. <https://doi.org/10.3758/BF03199010>
- R Core Team (2020) R: A Language and Environment for Statistical Computing (Version 2020) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reynoso-Cruz JE, Hernández-Salazar LT, Nieto J (2021) Forgetting curve in spider monkeys (*Ateles geoffroyi*): effects of delay but not the amount of information. *J Comp Psychol* 135(4):468–479. <https://doi.org/10.1037/com0000284>
- Rosenzweig ES, Barnes CA (2003) Impact of aging on hippocampal function: plasticity, network dynamics, and cognition. *Prog Neurobiol* 69(3):143–179. [https://doi.org/10.1016/S0301-0082\(02\)00126-0](https://doi.org/10.1016/S0301-0082(02)00126-0)
- Salthouse TA (2019) Trajectories of normal cognitive aging. *Psychol Aging* 34(1):17–24. <https://doi.org/10.1037/pag0000288>
- Schiel N, Souto A (2017) The common marmoset: An overview of its natural history, ecology and behavior. *Dev Neurobiol* 77(3):244–262. <https://doi.org/10.1002/dneu.22458>
- Schubiger MN, Wüstholtz FL, Wunder A, Burkart JM (2015) High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (*Callithrix jacchus*). *Anim Cogn* 18(3):701–712. <https://doi.org/10.1007/s10071-015-0837-5>
- Schubiger MN, Kissling A, Burkart JM (2016) How task format affects cognitive performance: a memory test with two species of New World monkeys. *Anim Behav* 121:33–39. <https://doi.org/10.1016/j.anbehav.2016.08.005>
- Schubiger MN, Fichtel C, Burkart JM (2020) Validity of Cognitive tests for non-human animals: pitfalls and prospects. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2020.01835>
- Sehner S, Willems EP, Vinicus L, Migliano AB, van Schaik CP, Burkart JM (2022) Problem-solving in groups of common marmosets (*Callithrix jacchus*): more than the sum of its parts. *PNAS Nexus* 1(4):pgac168. <https://doi.org/10.1093/pnasnexus/pgac168>
- Solomon SG, Rosa MGP (2014) A simpler primate brain: the visual system of the marmoset monkey. *Front Neural Circuits*. <https://doi.org/10.3389/fncir.2014.00096>
- Souto A, Bezerra BM, Schiel N, Huber L (2007) Saltatory search in free-living *Callithrix jacchus*: environmental and age influences. *Int J Primatol* 28(4):881–893. <https://doi.org/10.1007/s10764-007-9165-1>
- van Schaik CP, Jacobs I, Burkart JM, Sauciuc G-A, Schuppli C, Persson T, Song Z (2024) Short-term memory, attentional control and brain size in primates. *R Soc Open Sci* 11(5):231541. <https://doi.org/10.1098/rsos.231541>
- Vannuchi CRS, Costa CS, de Jesus FM, Maior RS, Barros M (2020) Sex, diurnal variation and retention interval differently affect performance of marmoset monkeys in a recognition memory task for object location. *Behav Brain Res* 379:112334. <https://doi.org/10.1016/j.bbr.2019.112334>
- Ward A, Webster M (2016) Sociality. In: Ward A, Webster M (eds) *Sociality: the behaviour of group-living animals*. Springer, pp 1–8. [https://doi.org/10.1007/978-3-319-28585-6\\_1](https://doi.org/10.1007/978-3-319-28585-6_1)
- Xavier DP, Abreu F, Souto A, Schiel N (2024) Choosing the best way: how wild common marmosets travel to efficiently exploit resources. *Anim Cogn* 27(1):20. <https://doi.org/10.1007/s10071-024-01864-8>
- Yamamoto ME, Araújo A, de Sousa MBC, de Arruda MF (2010) Social organization in *Callithrix jacchus*. *Advances in the study of behavior*, vol 42. Elsevier, pp 259–273. [https://doi.org/10.1016/S0065-3454\(10\)42008-2](https://doi.org/10.1016/S0065-3454(10)42008-2)
- Yamazaki Y, Saiki M, Inada M, Watanabe S, Iriki A (2016) Sustained performance by common marmosets in a delayed matching to position task with variable stimulus presentations. *Behav Brain Res* 297:277–284. <https://doi.org/10.1016/j.bbr.2015.10.025>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.