The Evolution of Primate Short-Term Memory

ManyPrimates[*], Géraud Aguenounon[1,2], Matthias Allritz[3], Drew M. Altschul[4], Sébastien Ballesta[1,2], Alice Beaud[1,2], Manuel Bohn[*,5], Sally L. Bornbusch[6], Angela Brandão[7,8], James Brooks[9], Thomas Bugnyar[10], Judith M. Burkart[11], Léa Bustamante[1,2], Josep Call[4], Charlotte Canteloup[12,13], Chuangshi Cao[14], Kai R. Caspar[15], Diana da Silva[16], Alexandra A. de Sousa[17], Sarah E. DeTroy[5], Shona Duguid[3], Timothy M. Eppley[18,19], Claudia Fichtel[20], Julia Fischer[21], Chi Gong[14], James A. Grange[22], Nicholas M. Grebe[6], Daniel Hanus[5], Daniel Haun[5], Lou M. Haux[21], Yseult Héjja-Brichard[24], Annabella Helman[6], Istvan Hernadi[25], R. Adriana Hernandez-Aguilar[26,27], Esther Herrmann[28], Lydia M. Hopper[29,30], Lauren H. Howard[31], Lei Huang[14], Sarah M. Huskisson[29], Ivo Jacobs[32,33], Zhiyong Jin[34,35], Marine Joly[26], Fumihiro Kano[36,37], Stefanie Keupp[21], Evelin Kiefer[25], Balázs Knakker[25], Katalin Kóczán[25], Larissa Kraus[5], Sze Chai Kwok[34,35,38,39], Marie Lefrançois[1,2], Laura Lewis[40], Siyi Liu[41], Miquel Llorente[42], Elizabeth Lonsdorf[43,44], Louise Loyant[28], Katarzyna Majecka[45], Luke Maurits[5], Hélène Meunier[1,2], Flávia Mobili[46], Luca Morino[47], Alba Motes-Rodrigo[48], Vincent Nijman[49], Caroline Nkov Ihomi[1,2], Tomas Persson[32,33], Dariusz Pietraszewski[50], Juan Felipe Reátiga Parrish[1,2], Anthony Roig[2,51], Alejandro Sánchez-Amaro[5], Yutaro Sato[9], Gabriela-Alina Sauciuc[32,33], Allie E. Schrock[6], Manon K. Schweinfurth[3], Amanda Seed[3], Caroline L. Shearer[6], Vedrana Šlipogor[10,52], Yanjie Su[41], Kirsten Sutherland[5], Jingzhi Tan[53], Derry Taylor[54], Camille A. Troisi[55], Christoph J. Völter[3,56], Elizabeth Warren[57], Julia Watzek, and Pauline Zablocki-Thomas[58]

[1]Laboratoire de Neurosciences Cognitives et Adaptatives, Université de Strasbourg; [2]Centre de Primatologie de l'Université de Strasbourg; [3]School of Psychology and Neuroscience, University of St Andrews; [4]Department of Psychology, The University of Edinburgh; [5]Max Planck Institute for Evolutionary Anthropology; [6]Department of Evolutionary Anthropology, Duke University; [7]Animal Behavior Department, University of Lisbon; [8]Center for Philosophy of Sciences, University of Lisbon; [9]Wildlife Research Center, Kyoto University; [10]Department of Behavioral and Cognitive Biology, University of Vienna; [11]Department of Anthropology, University of Zurich; [12]Inkawu Vervet Project, South Africa; [13]University of Lausanne; [14]Wuhan Zoo, Wuhan, China; [15]Department of General Zoology, University of Duisburg-Essen; [16]Primate Cognition Research Group, Center for Philosophy of Sciences, University of Lisbon; [17]Bath Spa University; [18]Conservation Science & Wildlife Health, San Diego Zoo Wildlife Alliance; [19]Department of Anthropology, Portland State University; [20]Behavioral Ecology & Sociobiology Unit, German Primate Center - Leibniz Institute for Primate Research; [21]Cognitive Ethology Laboratory, German Primate Center - Leibniz Institute for Primate Research; [22]School of Psychology, Keele University; [23]Center for Adaptive Rationality, Max Planck Institute for Human Development; [24]Centre d'Ecologie Fonctionnelle et Evolutive, Université de Montpellier, CNRS, EPHE, IRD, Montpellier; [25]UP-Gedeon Richter Grastyan Translational Research Center, University of Pecs; [26]Department of Social Psychology and Quantitative Psychology, Serra Hunter Programme, University of Barcelona; [27]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo; [28]Centre for Comparative and Evolutionary Psychology, Department of Psychology, University of Portsmouth; [29]Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo; [30]Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine; [31]Department of Psychology, Scientific and Philosophical Studies of the Mind, Franklin & Marshall College; [32]Department of Philosophy / Cognitive Science, Lund University; [33]Lund University Primate Research Station; [34]Phylo-Cognition Laboratory, Division of Natural and Applied Sciences, Duke Kunshan University, Duke Institute for Brain Sciences; [35]Shanghai Key Laboratory of Brain Functional Genomics, School of Psychology and Cognitive Science, East China Normal University; [36]University of Konstanz; [37]Kumamoto Sanctuary, Kumamoto, Japan; [38]Shanghai Key Laboratory of Magnetic Resonance, East China Normal University; [39]Shanghai Changning Mental Health Center, Shanghai, China; [40]Department of Human Evolutionary Biology, Harvard University; [41]School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Peking University; [42]Serra Húnter Fellow, Department of Psychology, University of Girona; [43]Department of Psychology, Biological Foundations of Behavior Program, Franklin & Marshall College; [44]Department of Anthropology, Emory University; [45]The Natural History Museum, Faculty of Biology and Environmental Protection, University of Lodz; [46]Department of Anthropology, University of Zurich; [47]Muséum National d'Histoire Naturelle, Paris, France; [48]Department of Ecology and Evolution, University of Lausanne; [49]School of Social Science, Oxford Brookes University; [50]Department of Ecology and

Vertebrate Zoology, Faculty of Biology and Environmental Protection, University of Lodz; [51]Normandie University; [52]Department of Zoology, Faculty of Science, University of South Bohemia; [53]Institute of Psychology, Leiden University; [54]Institute of Biology, University of Neuchâtel; [55]Department of Experimental Psychology, Ghent University; [56]Comparative Cognition, Messerli Research Institute, University of Veterinary Medicine Vienna; [57]University of St. Andrews; [58]California National Primate Research Center

*Corresponding author (Email: manyprimates@gmail.com)

Abstract – Short-term memory is implicated in a range of cognitive abilities and is critical for understanding primate cognitive evolution. To investigate the effects of phylogeny, ecology and sociality on short-term memory, we tested the largest and most diverse primate sample to date (421 non-human primates across 41 species) in an experimental delayed-response task. Our results confirm previous findings that longer delays decrease memory performance across species and taxa. Our analyses demonstrate a considerable contribution of phylogeny over ecological and social factors on the distribution of short-term memory performance in primates; closely related species had more similar short-term memory abilities. Overall, individuals in the branch of *Hominoidea* performed better compared to *Cercopithecoidea*, who in turn performed above *Platyrrhini* and *Strepsirrhini*. Interdependencies between phylogeny and socioecology of a given species presented an obstacle to disentangling the effects of each of these factors on the evolution of short-term memory capacity. However, this study offers an important step forward in understanding the interspecies and individual variation in short-term memory ability by providing the first phylogenetic reconstruction of this trait's evolutionary history. The dataset constitutes a unique resource for studying the evolution of primate cognition and the role of short-term memory in other cognitive abilities.

Keywords – Cognitive evolution, Short-term memory, Primate cognition, Phylogenetic analysis

_____

Inferring *how*, *why*, and *when* psychological traits evolved is a goal central to the field of comparative cognition (MacLean et al., 2012). The cognitive capacities of nonhuman primates ("primates" hereafter) have particular relevance for understanding the emergence of human cognition (MacLean, 2016). A systematic, comparative analysis of behaviors in extant primates is the best tool available to infer cognitive characters in ancestral taxa, and accordingly estimate when and in which lineages particular cognitive traits emerged and how they changed over the course of evolution (Shettleworth, 2010). Compared to many other mammals, primates have evolved enlarged brains relative to body size (Boddy et al., 2012) and higher neuronal densities (Herculano-Houzel et al., 2007). Consistent with this remarkable neural architecture, many primate species have been found to display advanced skills in various domains such as memory (Ghazizadeh et al., 2018; Lewis et al., 2019), tool making and use (Resende et al., 2021; Sanz et al., 2013; Shumaker et al., 2011; Wynn et al., 2011), planning (Osvath, 2009; Prétôt & Brosnan, 2019), causal reasoning (Cacchione & Rakoczy, 2017; Völter & Call, 2017), theory of mind (Crockford et al., 2012; Krupenye & Call, 2019), and metacognition (Basile et al., 2015; Beran et al., 2015; Rosati & Santos, 2016).

What social and ecological forces drive the evolution of these complex cognitive traits in primates? Popular hypotheses have highlighted the particular demands associated with facing ecological variation and living in a dynamic social environment. The Ecological Intelligence Hypothesis (Milton, 1981; Rosati, 2017) posits that living in a structurally complex habitat partly drives the evolution of specialized and

advanced cognitive skills in primates, for reasons including: 1) memorizing numerous locations and anticipating when key food resources are available is an important fitness advantage for primates (Milton, 1981); 2) living in an unpredictable environment requires a certain degree of cognitive flexibility to regularly discover new resources or develop new behavioral strategies in order to cover minimal maintenance costs (Sol, 2009); and 3) developing technical skills (e.g., extractive foraging, tool use) offers the opportunity to get access to exclusive and/or calorie-rich resources (e.g., nuts or social insects) (Byrne, 1997).

The Social Intelligence Hypothesis, and derived variations thereof, argues that group-living and the subsequent need to compete and/or cooperate with conspecifics are key drivers shaping the evolution of primate cognition (Burkart et al., 2009; Byrne & Whiten, 1988; Dunbar, 1998; Dunbar & Shultz, 2007; Jolly, 1966; Tomasello & Call, 1997). Most primates are social and live in groups ranging from two to hundreds of individuals (Smuts et al., 1986). Being able to track individual relationships and recognize allies and competitors is thus crucial to foraging, mating, and rearing offspring. Note that the two hypotheses are not mutually exclusive, specifically in the context of the competition over food.

Hypotheses regarding the drivers of cognitive and brain evolution can be fruitfully tested by comparing different primate species with respect to their cognitive abilities. To succeed, this approach requires datasets that span a large number of species living in different ecological and social environments. Primates are well-suited to such an inquiry by virtue of their diverse diets (ranging from frugivory, to folivory and omnivory; DeCasien et al., 2017) and social systems (ranging from group living, to pair living, and more solitary lifestyles; Kappeler, 1999). However, such datasets are not readily available due to the logistic difficulties associated with testing large numbers of primates and, accordingly, little empirical research has explicitly tested these hypotheses (Hopper et al., 2018). Indeed, a recent review found that only a fraction of comparative primate cognition studies include more than two species (ManyPrimates, 2019a). Collaborative, multi-site research efforts with multiple species are thus required to address these types of evolutionary and phylogenetic questions. In this study, we leveraged a previously established infrastructure (ManyPrimates, 2019b, see also below) and tested one of the largest and most diverse datasets ever collected using a standardized experimental procedure. We investigated the evolution of one of the most fundamental cognitive abilities in primates: short-term memory.

The concept of short-term memory (STM)—the ability to hold active representations over short periods of time—has played a central role in psychological research for centuries and can be traced to the very roots of modern psychology (Atkinson & Shiffrin, 1971). STM is usually contrasted with working memory, which captures the ability to mentally manipulate such representations (Cowan, 2008). Neuroscientific studies in humans suggest that STM is also distinct from other memory capacities such as long-term memory (Vallar & Papagno, 2002). It has been argued that STM is implicated in almost all cognitive tasks (Jonides et al., 2008). For instance, success on object permanence tasks requires remembering the location of hidden rewards after short delays (Zewald & Jacobs, 2021). From an evolutionary perspective, researchers have suggested that constraints on STM could explain species differences in cognitive abilities, such as a lack of recursive communication and complex tool-use in many primate species (Read, 2008). As such, understanding the origins of STM and its phylogenetic distribution represents an important goal for comparative psychology.

One key source of STM failure (i.e., forgetting) stems from time-based decay of mental representations. Thus, one valuable methodological tool for evaluating STM capacity is by varying retention intervals in a task (Barth & Call, 2006; Harlow et al., 1932; Maslow & Harlow, 1932; Nissen et al., 1938). For example, Mercer and McKeown (2014) presented (human) participants with two tones that they had to compare and implemented several conditions varying the interval of time between the presentation of the tones. The authors observed a significant drop in performance as the interval length increased. In the context of comparative cognition studies, retention interval often refers to the length of time between seeing an object hidden in a specific location and the subsequent testing of memory for the object's location. Empirical studies on STM in humans show that, during the retention interval of STM tasks, areas of the prefrontal cortex show elevated activity levels, suggesting that STM representations are

supported by neural activity in these areas (Grimault et al., 2009; Postle, 2015; Riley & Constantinidis, 2016).

Since the initial study by Hunter (1913) on "delayed reactions," the study of STM in primates has largely focused on only a few species, especially rhesus macaques (*Macaca mulatta*) (but see Harlow et al., 1932; Harlow & Bromer, 1939). STM in rhesus macaques is underpinned by neural mechanisms similar to those observed in humans (Constantinidis & Procyk, 2004) and is distinct from other memory capacities such as long-term memory (Zola-Morgan & Squire, 1986). While rhesus macaques' performance on STM tasks is significantly lower than human performance, there is evidence that patterns of primate STM parallel those in humans; for example, items presented early or late in a test set are better recalled than those in the middle (the "serial position" effect; Sands & Wright, 1980). Additionally, research with chimpanzees (*Pan troglodytes*), suggests that their STM is constrained in ways comparable to the 'magic number seven' (plus or minus 2) effect in humans, wherein humans often struggle in STM tasks that require subjects to recall more than seven items (Miller, 1956). In research using touch-screen technology to order numbers that became hidden after a short delay, chimpanzees' STM capacity was found to be between 4 and 9 items (Inoue & Matsuzawa, 2009; Kawai & Matsuzawa, 2000).

It is clear that primate species vary in their STM performance. However, due to the wide variety of methods used to study this phenomenon, it is difficult to systematically assess cross-species similarities and differences in STM from the published literature. Additionally, due to the fact that only 15% of the approximately 500 primate species have been included in studies of primate cognition, and only 19% of recent studies have included more than one species (ManyPrimates et al., 2019a), previous work has not been able to answer questions about evolutionary processes that account for species differences in STM. To address these limitations, we aimed to use a single test method to evaluate STM capacity across a wide range of primate species and provide the first phylogenetic reconstruction of the evolution of this trait.

Our study used the infrastructure established by the ManyPrimates project—a global consortium of researchers and study sites in primate cognition research. The main goal of ManyPrimates is to approach questions about the variability and evolution of primate cognition by collaboratively building datasets that include a wide range of species and individuals. ManyPrimates has established a sustainable and long-lasting infrastructure that enables collaboration between researchers and institutions around the globe. For more details about our project, we refer to previous publications (*cf.* ManyPrimates et al., 2019a, b, 2020, in press) and our website https://manyprimates.github.io.

In the present study, we systematically examined the evolution of primate visuospatial STM abilities by testing 421 individuals of 41 primate species across 29 sites (see Figure 1) in a delayed response task (see also ManyPrimates et al., 2019a). The diversity of individuals and species included in our sample allowed us to evaluate how individual and species-specific factors relate to STM. In the task, an individual watched an experimenter hiding a reward under one of three cups. The individual could then retrieve the reward by choosing the correct hiding location, and we examined their relative success rates across different retention intervals (i.e., three delay conditions: 0 s, 15 s, and 30 s). This study is a continuation of our pilot study (ManyPrimates et al., 2019a), in which we validated the methodology and generated hypotheses about the phylogenetic distribution of STM abilities. However, with only 12 species studied, the pilot study did not allow us to systematically investigate the processes underlying the evolution of primate STM abilities. Here, we included these earlier data but substantially extended it by testing a much larger number of subjects from a much broader range of species. With this much larger data set, the present study alleviates this shortcoming.

Figure 1

Overview of Data Collection Sites (Left) and Examples for Task Implementation (right)

Note. A) the 29 sites that contributed data to the study. B) pictures of the implementation of the delayed-response task for six species from six data collection sites (row-wise starting in upper left: *Macaca mulatta, Sapajus apella, Pongo abelii, Lemur catta, Gorilla gorilla gorilla, Macaca fascicularis*). For a video example of the task see: https://osf.io/mntpe/.

As a first step, we evaluated the effect of delay length on STM. Based on previous work (Barth & Call, 2006; Bartus et al., 1978) and on pilot data (ManyPrimates et al. 2019a) we predicted that species would differ in their performance on the task (Harlow et al. 1932) and that longer delays would lead to decreased performance across species and taxa. In this analysis, we also tested the influence of a range of individual-level predictors on performance, such as age or experience with delayed-response tasks. Based on previous studies, we predicted that task success would decrease with age (Elmore & Wright, 2015) whereas previous experience would lead to better task performance. In a second step, we used our experimental data to model the phylogenetic history of STM abilities. Based on pilot data and preliminary analysis (ManyPrimates et al., 2019a), we predicted that more closely related species would present more similar performance in the delayed response task. We then extend previous work by evaluating the relationship between socio-ecological factors and cognition (e.g., Cunningham & Janson, 2007; Rosati, 2017; Schwartz, 2019). We tested different hypotheses about how sociality and ecology contribute to the evolution of short-term memory abilities in primates, over and above phylogenetic relatedness. These hypotheses were solicited from the broader research community. We circulated a description of the expected dataset via social media and mailing lists to encourage researchers in the field to submit hypotheses specifying which social and ecological variables predict STM abilities across primates. These hypotheses were collected and translated into statistical models which we then compared in a phylogenetic model comparison. This approach allowed us to be as inclusive as possible and cover a broad range of theoretical positions. The submitted hypotheses included a wide range of predictors such as diurnal-resting time, trichromacy, home range, vocal repertoire size, dietary diversity, group size, day journey length, arboreality and frugivory.

Methods

Ethics Statement

Experiments and participating institutions complied with the ethics guidelines of the ManyPrimates project (https://manyprimates.github.io/ethics/) and explicit ethical approval was obtained

from each participating institution. In the Supplementary Material we provide a detailed description of each data collection site including housing and research practices and the procedure that was used to obtain ethical approval.

Subjects

We collected data from 421 captive primates, representing 41 different species of *Platyrrhini*, *Strepsirrhini*, *Cercopithecoidea*, and *Hominoidea* (nomenclature applied following Mittermeier et al., 2013). Subjects originated from 29 different sites located in 13 countries across the world that included zoos, sanctuaries, and laboratory facilities (Figure 1). A subset of this data has been published in ManyPrimates et al. (2019a) (see Table S1 in Supplementary Material).

We tested two families of lemurs: *lemurids* (true lemurs and relatives, 8 species, n = 72) and *indriids* (sifakas and relatives, 1 species, n = 9). We included three families of *Platyrrhini*: *cebids* (capuchin and squirrel monkeys, 3 species, n = 92), *callitrichids* (marmosets and tamarins, 4 species, n = 50) and *atelids* (spider monkeys and relatives, 2 species, n = 4). The majority of species and individuals studied represented the *Cercopithecidae* clade (*Cercopithecidae*). We studied both subfamilies of this group: the generalist *cercopithecines* (cheek-pouched monkeys, 10 species, n = 93 (including 2 hybrid subjects, *P. anubis* x *cynocephalus*)) and the *folivorous colobines* (leaf-eating monkeys, 3 species, n = 13). Finally, both families of apes were tested: the *hylobatids* (gibbons, 5 species, n = 9) and the *hominids* (great apes, 5 species, n = 79).

Our sample thus covers the majority of radiations within the primate order. However, Tarsiiformes (tarsiers) and *Lorisiformes* (lorises, galagos, and relatives) were not studied due to a lack of participating sites offering access to these animals. Several families of *Platyrrhini* (*Aotidae* and *Pitheciidae*) and lemurs (*Cheirogaleidae*, *Daubentoniidae*, *Lepilemuridae*) were not included for the same reason.

Materials

We tested all individuals in the delayed-response task. This paradigm was chosen given its internal validity, simplicity, and popularity in the animal cognition literature. The general setup comprised a rectangular board and three identical, opaque cups. High-value food items were used as rewards (Figure 1B; the size and food type varied across species and sites). In the test, the board was placed in front of the subject, outside the enclosure. The cups were evenly spaced on the board with at least 10 cm between them (center to center). The following aspects of the setup varied between sites: board size, cup size and color, distance between cups, food reward used, testing arrangements (group vs. individual testing), and subject experience with object choice tasks (task experience also varied between subjects at the same site). Some of these differences were due to differences between testing facilities (e.g., the testing arrangements), others due to differences between primate species or cohorts (e.g., the type and size of the food rewards). We documented these differences in the Supplementary Material (see Table S1) and considered some of these differences as predictors in our confirmatory analysis (board size, cup distance, task experience). A short video of the task with different species from different sites can be found here: https://osf.io/mntpe.

In our pilot study (ManyPrimates et al., 2019a), some aspects of the setup systematically co-varied with species (board size and cup distance). We therefore asked institutions to vary board size and cup distance independent of species body size (but given that we included our pilot data we could not completely remove this confound in this way). To ensure the proper implementation of the setup and the procedure, each site recorded a short video of the test setup and implementation of a trial prior to actual testing. The project coordinators checked the videos and provided feedback if they noticed any deviations from the protocol.

Design

The main experimental (within-subject) manipulation was the time passed between hiding a food item in full view of the subject and allowing the subject to retrieve it. The intervals were 0 s, 15 s, and 30 s resulting in three delay conditions: short, medium, and long.

Procedure

Depending on the site, the subjects were tested individually or in group settings. In the case of group settings, researchers specified the focal individual ahead of time and distracted non-focal individuals during testing by giving them access to additional food or enrichment items. Before participating in the test, some subjects with little object-choice experience received additional training to ensure reliable choice behaviors. They were trained to reliably point to or reach for fully visible food items placed on the board before the start of the experiment. In our choice training protocol, subjects were presented first just with an out-of-reach food item, then with a food item that was covered by an opaque cup. Finally, they had to choose between two cups, one visibly baited and the other one not. Only when subjects made reliable, unambiguous choices in each training phase were they presented with the test (but there was no formal progression criterion).

The number of trials per test day (session) varied between sites, species, and individuals, with the constraint that there were at least three trials (one block) per test day. To be included in the analysis, individuals had to contribute a minimum of 9 trials and a maximum of 36 trials, equally distributed across conditions.

At the beginning of a trial, the experimenter (E; familiarity between E and the subjects varied between sites) pulled the board back so that the subject could not reach it. E stood or sat behind the board and placed the three cups next to each other on their sides, with the opening facing the subject. Next, E showed the subject a food item and placed it in front of one of the cups. Then E put the cups down one by one, thereby hiding the food item, always starting with the cup on the left from E's perspective. Depending on the condition, E waited either 0 s (short delay), 15 s (medium delay) or 30 s (long delay) before pushing the board towards the subject. The delay started once E had put down the last cup (on the right). While pushing the board forward, E looked down and center to avoid inadvertent cueing. The subject made a choice by either pointing to or touching one of the cups. If the subject chose more than one cup simultaneously, E pulled the board back and pushed it forward again to ensure an unambiguous choice. If the subject did not make a choice within 60 s, the entire trial (including hiding and delay) was repeated. After the subject chose a cup, E pulled the board back and lifted the indicated cup. If the cup revealed the food item, the subject got it as a reward. If the cup did not cover the food item, the subject got no reward. After E had lifted the indicated cup (and passed the reward to the subject in case of a correct choice), E turned over the remaining two cups with the open side facing to the subject and, in case of an incorrect choice, took the remaining food reward back in preparation for the next trial. The same food item was used again in the next trial. Sessions were terminated if a subject did not make a choice in three consecutive trials. Further, data collection was stopped with a subject if three sessions had to be terminated because the subject did not make a choice.

For each subject, the hiding location was pseudo-randomized across trials with the constraints that the same location occurred no more than two times in a row and that each hiding location occurred an equal number of times per condition. Trials were grouped in blocks, with each block comprising three trials of the same condition (either short, medium, or long delay). Each hiding location occurred once within each block. Each set of nine trials comprised three blocks, one per condition. The order of conditions across the three blocks was randomized. Different subjects received different randomizations.

Scoring

We scored whether the subjects chose the correct cup, i.e., whether they manually indicated the location in which the food item was hidden. This resulted in a binary (0: incorrect, 1: correct) variable for each trial. Trials were filmed whenever possible. To assess inter-rater reliability of choice scoring, an independent coder re-coded at least 20% of the trials at each site. Table S1 in the supplementary material gives Cohen's Kappa values for each site (range: 0.72 - 1).

Analysis

The confirmatory and phylogenetic analyses were pre-registered before we started to inspect the data (https://osf.io/sf3bx). All data and analysis scripts are available in a public repository (https://github.com/ManyPrimates/mp1_short_term_memory). All statistical analyses were performed in R version 4.0.3 (R Core Team, 2020).

For the comparison to chance level, we aggregated the data for each individual in each of the three delay conditions and used the function ttestBF from the package BayesFactor (Morey & Rouder, 2018) to compute the Bayes Factor (based on a Bayesian t-test) in favor of the hypothesis that the average proportion of correct responses in a condition was above 0.33.

All other models were fitted to the trial-by-trial data as Bayesian generalized linear mixed models (GLMM) with a logit link using the function brm from the package brms (Bürkner, 2017). Model parameters for each model were estimated by collecting 10 000 samples from eight independent MCMC chains, removing the first 5,000 samples for burn-in. Priors for all models are reported in the Supplementary Material. All models converged without problems with Rhat values < 1.01. The model outputs can be accessed via the online repository.

We computed WAIC (widely applicable information criterion) scores for every model, and for the model comparison, we also added WAIC weights. Following McElreath (2018), we used these metrics to rank models. In addition, we inspected the 95% Credible Intervals for the test predictors of interest. The confirmatory model had the following structure:

```
correct ~ delay +

    task_experience + norm_age + cup_distance + board_size +

    trial +

    (1 + delay + trial | subject_site) +

    (1 + delay + trial | site) +

    (1 + delay + trial | species)
```

where `correct` noted whether an individual chose the correct cup on a given trial. `Delay` was the length of the time between hiding the food and choosing a cup. For the confirmatory analysis, this was coded as a three-level factor (centered at medium delay of 15 s). For the phylogenetic analyses, it was coded as a two-level factor (medium and long delay coded as 0 and short delay coded as 1). This factor coding is different from the numeric coding of delay we pre-registered. In the Supplementary Material, we explain why we deviate from our pre-registration here. `Task_experience` noted whether the individual has participated in comparable object-choice studies before (coded as yes/no). `Norm_age` was the individual's age normalized by the maximum recorded life span of that species. `Cup_distance` was the distance between

the cups and `board_size` the width of the board on which cups were presented (both in cm). `Trial` noted the trial continuously across sessions[1]. `Subject_site` was a unique identifier for each subject. `Site` was the data collection site and `species` noted the individual's species. All numerical predictors were scaled to have a mean of 0 and a standard deviation of 1. The alternative model in the confirmatory analysis had the same structure except that it did not include delay as a predictor.

The phylogenetic baseline model had the following structure:

```
correct ~ delay +

    (1 + delay + trial | subject_site) +

    (1 + delay + trial | site) +

    (1 + delay + trial | gr(species, cov=vcv.phylo(tree))
```

Here, `tree` is a consensus primate phylogeny from the 10ktrees project (Arnold et al., 2010), pruned to include only those species included in our study. The `vcv.phylo` function of the R package `ape` (Paradis & Schliep, 2019) was used to compute the expected correlations between the tip values (i.e., per-species values) of a trait evolving along this tree according to a Gaussian process model of evolution (including the Brownian motion model). The resulting matrix is used as the covariance matrix for all random effects of species.

In contrast to the confirmatory model, the phylogenetic models did not include the control predictors `board_size`, `cup_distance`, `task_experience,` and `trial`. The reasons were: a) some of these variables were confounded with phylogeny (e.g., `board_size` and phylogeny both co-varied with body size), b) we had no hypothesis about if and how they interact with the test predictors, and c) excluding them facilitated the interpretation of the influence that test predictors had on performance.

To assess the degree to which performance in the task follows a pattern expected by the phylogenetic relatedness between species, we compared the phylogenetic baseline model to an identical model that made no assumptions about the correlations between species' performances (i.e., random effect for species: `(1 + delay + trial | species)`). We also quantified the phylogenetic signal in the data: Following a systematic comparison of indices (Münkemüller et al., 2012), we report Blomberg's K and Pagel's λ (Blomberg et al., 2003). The first of these (K) can be understood as the ratio between the observed degree of variation in the value of some variable across the tree about its mean and the expected degree of variation under a Brownian motion (BM) model of evolution for that variable; the latter (λ) is instead a scaling factor which, applied to the internal branch lengths of a tree, maximizes the likelihood of the data under the same BM model - values below 1 shrink branches, effectively reducing the effect of deep history. Despite being different indices, Blomberg's K and Pagel's λ share some key characteristics: their lower bound is zero (i.e., phylogenetic independence), and more substantial deviations from zero indicate stronger relations between per-species performance and phylogeny (i.e., a Brownian motion model of trait evolution). Values close to one suggest that per-species performance is distributed as expected by phylogeny. We computed Blomberg's K and Pagel's λ for the predictions of our phylogenetic baseline model (see below). We escaped the circularity of measuring the phylogenetic signal in a model which explicitly takes phylogeny into account via a partial constraint: while the estimated per-species random effects reflect phylogeny in their correlations, the actual *size* of these effects relative to either the fixed effects or other non-phylogenetic random effects (such as those of site or individual) is not strongly

---

[1] We included trial 'continuously across sessions' instead of 'within session' to be able to account for learning effects over the course of the study.

constrained and depends on the data. Thus, the phylogenetic signal in the overall model output can in principle be arbitrarily low if this is what fits the data well.

The phylogenetic model also accounted for the repeated testing of individuals and the nesting of individuals within sites. Thus, in order to get per-species estimates of the signal, we repeatedly sampled one random individual and one random site to represent each species. For each of these samples, we computed the phylogenetic signal in the linear predictor for the 3 possible values of `delay`. We collected 3,333 samples of individuals and sites, resulting in ~10,000 calculations of the phylogenetic signal in total, yielding the distribution described below.

To test which social and ecological species-level characteristics relate to STM performance, we conducted a phylogenetic model comparison. Given the myriad of species-level characteristics to consider, any particular selection largely hinges upon one's theoretical views concerning which social or ecological factors drive cognitive evolution. ManyPrimates, as well as the broader research community, is composed of scientists with a wide range of theoretical perspectives. To do justice to this plurality, we decided to solicit theories from the research community as part of a "modeling challenge". In February of 2020, before data collection had been completed, we circulated a rough description of the expected dataset via social media (Twitter) and mailing lists (International Primatological Society, Cognitive Science Society) and asked researchers to submit theories nominating the species-level characteristics they deemed most predictive of primate STM. Theories were restricted to include only characteristics external to animals; that is, characteristics that reflect a species' social or ecological environment. We deliberately excluded commonly used *internal* predictors of cognitive performance – e.g., brain size – because they do not specify any external pressures that require adapting to and therefore provide no answer to the question of why a given species evolved an ability.

In addition to detailing their model, we also asked researchers to submit sources for the data on the species-level predictors in their model. For many submissions, this information was missing or incomplete and we had to search for it ourselves. We did not find sources for all predictors for all species. As a consequence, we could not include all submitted models in the model comparison. We did not find data for many predictors for two species tested (Allen's swamp monkey (*Allenopithecus nigroviridis*) and Hamlyn's monkey (*Cercopithecus hamlyni*)), and so we decided to exclude data from these species from the phylogenetic analysis.

We would also like to note that the basis for some species-level predictor variables was very sparse. In many cases, we had to use secondary sources because the primary source was inaccessible. At other times, multiple sources were available but yielded substantially different estimates of species-level predictors. We used the following steps to decide which source to use in case we had more than one. First, we prioritized the source that provided data for the most species (to ensure comparability across species). To fill the remaining gaps, we used the source that provided the next most data for the remaining species, iteratively proceeding until we were no longer able to obtain further multi-species estimates. Finally, we used sources for individual species with the largest reported sample size to fill any remaining gaps. The associated online repository contains the final spreadsheet with the data and the sources we used for our analyses. We believe our repository provides the best possible estimates of these predictors given available data. At the same time, by making our dataset publicly available, we welcome refinement of our estimates via future studies and/or unpublished data.

The models that entered the phylogenetic model comparison were constructed by adding the submitted predictor variables to the phylogenetic baseline model as a main effect and a fixed effect interaction with `delay` (centered at 15 s). For example, one submission suggested that group size was related to short-term memory performance. The fixed effects structure of the corresponding model was therefore `group_size * delay`. We compared a total of 10 distinct models, including the baseline model. Table 1 lists all the individual predictors.

As part of the exploratory analysis, we fit additional main effects models for predictors submitted in models in which they were combined with others. For example, one submission included the term `percent_frugivory * terrestriality`. From this, we constructed two additional models, one

for `percent_frugivory` and the other for `terrestriality`. The results of this analysis are reported in the Supplementary Material.

Table 1

Predictors Used to Infer Short-term Memory (STM) Evolution in Primates

Predictor	Definition	Variable type
`body_size`	Average female body mass of a species (kg).	numeric
`color_vision`	Species' capacity to differentiate colors. Levels were dichromatic, trichromatic or polymorphic (i.e., both dichromatic and trichromatic individuals are present in a population).	factor (three levels)
`day_journey_length`	Daily traveled distance (in m)	numeric
`diet_diversity`	Trophic guild a species belongs to (e.g., folivore, frugivore, omnivore)	factor (six levels)
`dietary_breadth`	Number of dietary categories eaten by each species. Categories were defined as vertebrate, invertebrate, fruit, flowers/nectar/pollen, leaves/branches/bark, seeds, grass and roots/tubers.	numeric
`feeding_budget`	Percentage of time spent feeding from the total activity budget of a species.	numeric
`group_size`	Average number of individuals in social group	numeric
`home_range`	Average home range size (in ha).	numeric
`percent_frugivory`	Percentage of fruit in a species' diet.	numeric
`diurnal_resting_time`	Percentage of time spent resting during the day	numeric
`terrestriality`	Primary habitat (levels: arboreal or terrestrial)	factor (two levels)
`vocal_repertoire`	Number of distinct calls in a species' vocal repertoire	numeric

Note. Data sources for the different predictors are listed in the spreadsheet in the associated online repository.

In the Supplementary Material, we provide a number of additional analyses. First, we report a prior-sensitivity analysis for the phylogenetic models in which we constrained the influence of phylogeny on the model by reducing the variation between per-species random effects. We also report the results of several exploratory analyses, including an assessment of how variable species-level performance was across data collection sites, an assessment of the task's split-half reliability (which was acceptable at $r = 0.62$) as well as additional variants of the confirmatory and phylogenetic models.

Results

Descriptives

As a group, the primates we tested successfully remembered where food was hidden in the delayed-response task. When averaging across all individuals and species, the proportion of correct responses was higher than a level expected by random guessing (33% correct) in all conditions; $m_{short} = 0.76$ (SD = 0.24; Bayes Factor (BF$_{10}$) in favor of the hypothesis that performance was above $0.33 = 4.2 * 10^{112}$); $m_{medium} = 0.55$ (SD = 0.25; BF$_{10} = 2.7 * 10^{50}$); $m_{long} = 0.51$ (SD = 0.24; BF$_{10} = 6.5 * 10^{37}$); see also

Figure 2. When comparing across species, performance levels appeared to be clustered within clades and along phylogenetic lines: *Hominoidea* had a higher proportion of correct responses compared to *Cercopithecoidea*, who in turn performed above *Platyrrhini* and *Strepsirrhini* (Figure 2). However, performance varied greatly between individuals, such that within each delay condition, the response distribution for a given species overlapped with that of most other species.

Figure 2

Chronogram (Left) and Performance in the Delayed Response Task (Right) for the 41 Primate Species Tested
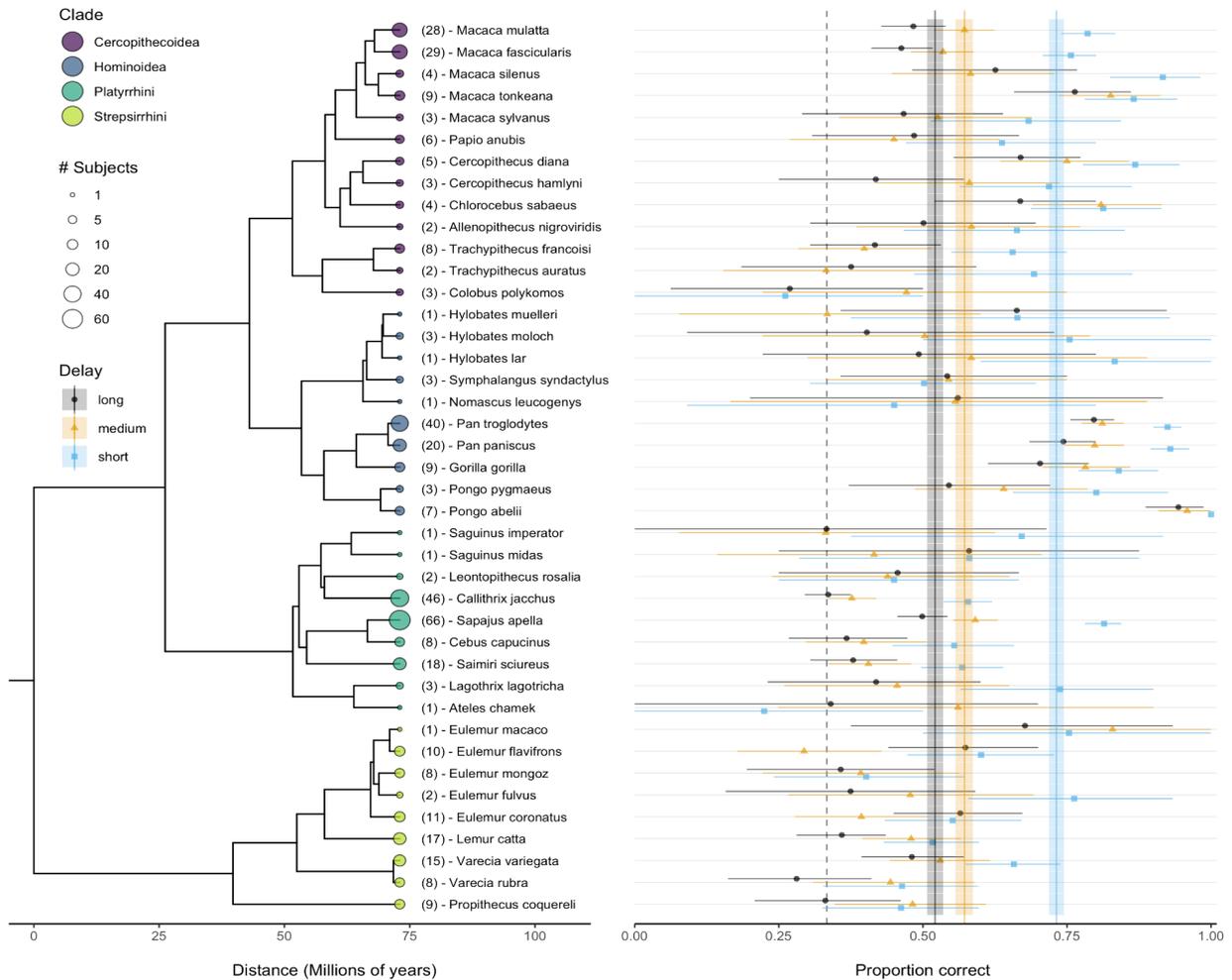
Note. Phylogenetic data were obtained from 10kTrees (Arnold et al., 2010). Branch lengths are proportional to absolute time. The size of the filled points is proportional to the number of subjects for each species. Colored shapes (with 95% CI) show the mean performance per species in the three delay conditions. The colored vertical lines show the mean performance across species (with 95% CI) in the three delay conditions. The dotted vertical line shows a level of performance expected by chance (33% correct).

Confirmatory Analysis

Our confirmatory analysis tested if delay length (i.e., retention interval) affected primates' performance in the delayed response task. To assess the effect of delay length, we compared a full model that included delay as a predictor to a model lacking it. A Bayesian model comparison based on WAIC (widely applicable information criterion) scores and weights clearly favored the model including delay

(delay: WAIC = 14205.50, se = 88.47, weight = 1.00; no delay: WAIC = 14786.99, se = 80.52, weight = 0.00). Compared to the medium (15 s) delay condition, performance was better in the short (0 s) delay condition ($\beta = 0.90$, 95% credible interval (CrI): 0.67 - 1.13) and worse in the long (30 s) delay condition ($\beta = -0.25$, 95% CrI: -0.36 - -0.12). Thus, performance in the delayed response task decreased when the delay between hiding the food and retrieving it increased, replicating the results of previous studies (Barth & Call, 2006; ManyPrimates et al., 2019a) and extending it to 29 new species (see Supplementary Material Table S1).

The confirmatory model also included a range of control predictors, which reflected variability between individuals and the physical implementation of the task. Figure 3 shows that individuals with experience in delayed response tasks performed better than individuals naïve to such tasks ($\beta = 0.25$; 95% CrI: 0.01 - 0.49). Individuals who were tested with cups further apart were also more likely to perform better ($\beta = 0.48$; 95% CrI: 0.26 - 0.71). The latter result is likely to be confounded with body size (and phylogeny); larger individuals (mostly great apes) were tested with cups placed further apart (see Supplementary Material).

Figure 3

Posterior Distributions for Predictors in the Confirmatory Analysis
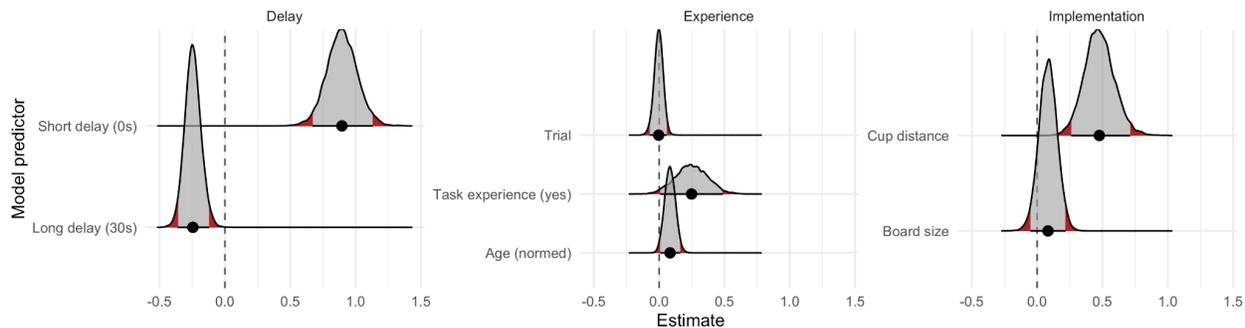
Note. Gray regions (and error bars) show 95% CrIs.

Performance increased with age (normalized by the maximal recorded lifespan for the species; $\beta = 0.08$; 95% CrI: 0.01 - 0.16), contradicting our pre-registered prediction that STM abilities would decline with age. In the Supplementary Material, we also report a model that posits a quadratic relation between age and performance to explore the possibility that performance increases early in life but then drops off again in old age. However, adding this nonlinear term did not improve the fit of the model[2].

Phylogenetic Analyses

Through a set of phylogenetic analyses, we tested whether performance in the delayed response task was linked to phylogeny and species-level predictors. The species in our sample differed in the amount of evolutionary history they shared; more closely related species might have more similar STM abilities because they evolved from a common ancestor. The confirmatory analyses reported above did not account for this potential source of structure in our data. As noted in the Methods section, for our phylogenetic analyses, we constrained the covariance between species-level effects to reflect the expected correlations between species based on a Brownian motion model of evolution applied to a consensus primate phylogeny (Arnold et al., 2010). When compared to a model that made no assumptions about the covariance between

---

[2] We did not pre-register the inclusion of testing situation (group vs. individual) as a predictor. This was because we used the same confirmatory model as for the pilot study, which did not include this predictor due to a lack of variation in the data. In a model that includes testing situation as a predictor in the confirmatory model, we find no difference between testing situations ($\beta = -0.03$, 95% CrI: -0.30 - 0.25) while all other predictor estimates remain unchanged.

species, a phylogenetic model provided a much better fit to the data (phylogeny: WAIC = 14091.57, se = 87.58, weight = 0.87; no phylogeny: WAIC = 14095.37, se = 87.53, weight = 0.13). This result suggests that species differences in performance in the delayed response task map onto phylogenetic relatedness between species (Figure 4).

Phylogenetic Signal

This result was also reflected in the indices quantifying the phylogenetic signal. For Blomberg's K the overall mean was 0.33 (95% HDI: 0.17 - 0.52) and for Pagel's λ the overall mean was 0.70 (95% HDI: 0.41 - 0.99). Thus, in both cases, the phylogenetic signal was measured to be reliably different from zero, suggesting a considerable overlap between the distribution of performance in the delayed-response task and species phylogeny.

Figure 4

Phylogenetic Tree with Ancestral State Reconstruction Based on a Brownian Motion Model of Trait Evolution for STM Performance (Left) and the Predictors of the Three Highest Ranking Models in the Model Comparison (Right)
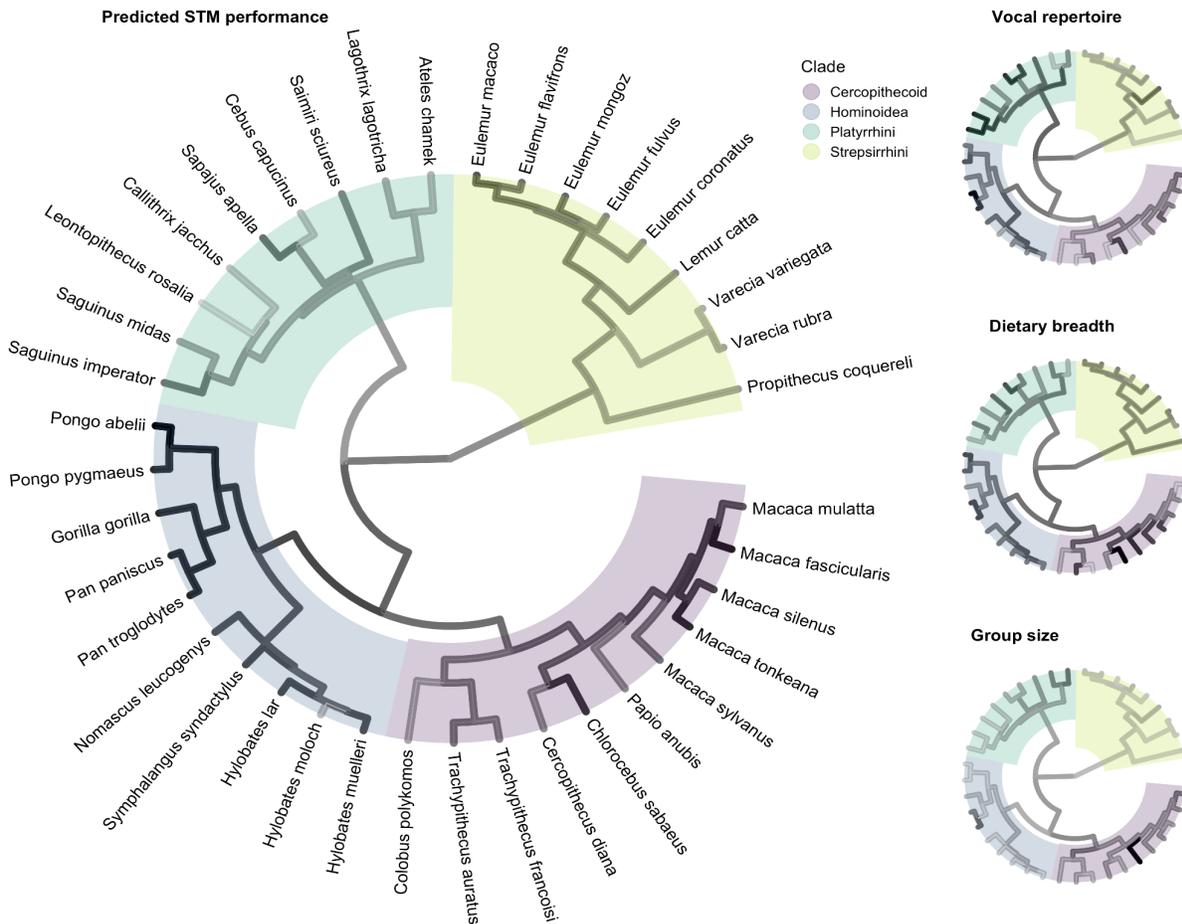
Note. For STM, darker lines correspond to higher probabilities (as predicted by the model) of choosing the correct cup in delay trials. For the predictors, darkness of the lines corresponds to the estimated predictor value, with darker lines corresponding to higher values (the scale differs across variables). Branch lengths are proportional to absolute time. Shading highlights the different clades.

The substantial phylogenetic signal in our dataset indicates that closely related species have more similar STM abilities. Still, phylogenetic relatedness cannot fully explain why some species have evolved more potent STM skills than others, as relatedness does not explain what non-genetic (e.g., environmental or social) factors have exerted selective pressure on different species. Thus, in the next analysis, we examined which species-level social or ecological characteristics were associated with STM abilities over and above phylogeny.

Phylogenetic Model Comparison

In the phylogenetic model comparison, we compared predictor models that specified one or more species-level, external variables that (presumably) relate to performance to a baseline mode that only accounted for phylogeny. Surprisingly, none of the predictor models outperformed the baseline model. All predictor models had WAIC scores higher than the baseline model and thus lower relative weight (see Table 2). However, the minute differences in WAIC scores and weights between the baseline, the vocal repertoire, and the dietary breadth model showed that these models were basically indistinguishable from one another. Because the baseline model offered a more parsimonious explanation of the data, we nonetheless conclude that the additional species-level predictors contributed little to explaining differences between species over and above phylogeny. In other words, knowing, for example, a species' vocal repertoire provides little additional information about STM abilities when already knowing its position in the phylogenetic tree.

Table 2

Phylogenetic Model Comparison

Model (predictors)	WAIC	SE	weight
(baseline)	14091.57	87.58	0.20
vocal_repertoire	14091.70	87.66	0.19
dietary_breadth	14092.19	87.67	0.15
group_size	14092.83	87.68	0.11
home_range/body_size + feeding_budget * dietary_breadth	14093.14	87.81	0.09
diurnal_resting_time	14093.39	87.67	0.08
percent_frugivory * terrestriality	14093.83	87.79	0.07
color_vision	14094.19	87.72	0.06
day_journey_length * group_size	14095.12	87.85	0.03
diet_diversity	14096.46	87.93	0.02

Note. All models had the same basic structure (baseline model): **correct ~ delay + (1 + delay + trial | subject_site + site + species)** with **species** capturing phylogeny following a Brownian Motion model of trait evolution (see main text for details). The predictors in the table were added to this baseline model as interactions with **delay**.

The distribution of the predictor variables across species may explain this result. We computed the phylogenetic signal in our predictor variables and found many to be strongly aligned with phylogeny (see Table S4 in the Supplementary Material). Figure 4 uses ancestral state reconstruction to visualize the phylogenetic distribution of STM performance as well as the predictor variables from the three highest-ranking models. While many of the predictor variables were positively correlated with STM abilities when ignoring phylogeny (see Figure S4 Supplementary Material), once accounting for phylogeny, these variables provide little additional information about species' STM abilities. Our results thus suggest that observed patterns arise because closely related species tend to live in more similar environments that share certain social and ecological features, rather than environmental features—at least the ones we investigated—independently affecting the evolution of STM abilities.

Discussion

In this study we compared short-term memory (STM) abilities across the largest and most diverse primate sample used in an experimental study to date. Primates from all clades performed above chance in at least one of the delay conditions, so that STM could be regarded as a basal cognitive trait of primates. Across species, we also found that the longer the delay between hiding and retrieving food rewards, the worse the performance on memorizing the location of the food. This finding confirmed our predictions and replicated the results of previous tests of nonhuman primates using the delay-response task (Barth & Call, 2006). A similar delay effect is also found in human studies of non-verbal auditory memory and is interpreted as reflecting the construal of STM as a transient storage of limited capacity (e.g., Mercer & McKeown, 2014).

The primates' performance was further influenced by phylogeny, age, and experience. Individuals with prior experience in such delayed response tasks performed better in this task than naïve individuals, and performance increased with relative age. A similar age effect has also been reported in human children, whose performance in a two-cup delayed-response task with increasingly long delays was positively predicted by the children's age (Diamond & Doar, 1989). Phylogenetic analyses revealed that species differences can be predicted by their phylogenetic relatedness. That is, closely related species performed more similarly in the delayed-response task compared to more distantly related species.

Within the phylogenetic tree, we found higher performance by individuals in the branch of *Hominoidea* compared to *Cercopithecoidea*, who in turn performed above *Platyrrhini* and *Strepsirrhini*. This pattern was not absolute in that, for example, some *Cercopithecoidea* (e.g., *Macaca tonkeana*) performed comparably to the *Hominidae* (e.g., *Gorilla gorilla*), while others (e.g., *Trachypithecus francoisi*) performed more similarly to some of the *Strepsirrhini* (e.g., *Lemur catta*). Additionally, performance varied greatly between individuals, such that the response distribution for a given species overlapped that of most other species within each delay condition, reflecting previous research in macaques and gorillas that has highlighted inter-individual variation in performance in tests of memory and learning (e.g., Altschul et al., 2016; Egelkamp et al., 2019).

Overall, however, the pattern of results we observed suggests systematic variation across species. This opens up the question of which selective pressures are responsible for the pattern we observed. We approached this question in our phylogenetic model comparison. The different models represent a range of hypotheses about which social or ecological variables may influence STM abilities. The models test the assumption that if one were to—hypothetically—intervene on the respective predictor variable(s), the STM abilities of a species should change over time - irrespective of the species' location in the phylogenetic tree. It is important to note that this approach is not able to detect effects of predictors if they are confounded with phylogenetic relatedness. However, this clustering would suggest that the predictor did not have a causal effect in and of itself, but rather that the effect was conditional on other factors that are only present in certain parts of the phylogenetic tree. From a statistical point of view, we were thus looking for predictors that explained species differences once phylogenetic relatedness had been accounted for.

Contrary to our predictions, we found no evidence that any of the predictor variable(s) we considered had such a general effect. None of the models that included a predictor substantially outperformed a baseline model which only accounted for phylogeny. That is, when knowing a species' location in the phylogenetic tree, learning about its social or ecological characteristics (e.g., vocal repertoire or dietary breadth) did not substantially improve the ability to predict that species' STM capacity. Thus, on the one hand the strong phylogenetic signal suggests some systematic pressures working on the evolution of STM abilities, but on the other hand we did not identify these pressures in our phylogenetic model comparison. These results contrasts with previous research that has found links between primate brain size (as a proxy for cognitive abilities) and some of the predictors also included here (e.g., diet and home range, DeCasien et al., 2017; Powell et al., 2017)

There are a number of (mutually non-exclusive) reasons why the predictors we considered did not contribute to explaining the phylogenetic distribution of STM abilities. First, the predictors that were submitted as part of the modeling challenge could truly have no influence and other species-level social or

ecological variables need to be considered. In a similar way, other cognitive abilities (such as other executive functions, like control of attention; Morey & Bieler, 2013) could be responsible for the differences across species in performance in the delayed-response task. Such a scenario assumes that variation in STM abilities in primates are essentially a product of other evolutionary processes which would shift the locus of explanation to identifying which variables predict performance in other cognitive abilities. Second, the effect of the predictor variables might not be as general as assumed in the models. For example, dietary breadth might have an effect on STM abilities only when considered in conjunction with home range size. Third, the predictors might be measured at the wrong time and for the wrong species. We estimated the predictors based on data from extant species, which leaves open the possibility that they might have been different at the time in which they exerted a selective pressure on ancestral species. Finally, the measurement quality of the predictors could be insufficient. In fact, in many cases, we had to rely on secondary sources to find values for some predictors. Thus, it might be that our predictor values do not accurately represent species' sociality or ecology. If this were the case, however, it would also cast doubt on previous work because we relied on published studies for most of our predictor values. Other reasons are of course conceivable. All of this shows that there is still much work to be done when it comes to identifying the selective pressures that shaped the evolution of STM. Nevertheless, the dataset we collected here provides a crucial resource to tackle this problem. Crucially, it also allows researchers to test additional hypotheses that were not considered as part of the modeling challenge.

The results of the present study also inform the debate around the neural mechanisms of STM. The dorso-lateral prefrontal cortex has been suggested to be a 'necessary and sufficient' substrate of STM (as suggested by e.g., Riley & Constantinidis, 2016). However, the successful performance of some lemurs in the present study, and the absence of a granular dorso-lateral prefrontal cortex in strepsirrhine species (Wise, 2017), speaks against this hypothesis. The emergence of a dorso-lateral prefrontal cortex in anthropoids may have contributed to a more robust STM through increased executive control (consistent with e.g., Postle, 2015). Consequently, STM efficiency would be further boosted in primates endowed with a more complex dorso-lateral prefrontal cortex, enabling individuals to withstand interference or mere time-based decay and, thus, to hold items in memory for longer periods of time. More research using additional STM tasks of various complexity and modalities are needed to adjudicate between these various possibilities.

An interesting extension to the current work would be to examine working memory across species using the ManyPrimates approach. While STM refers to the passive retention of information over the short-term, working memory is a multi-component system responsible for the active maintenance and manipulation of information (Cowan, 2008; see Basile & Hampton, 2013, for discussion of the differentiation between STM and working memory in non-human primates). Although models differ in their explanation of working memory, all share the requirement of attentional control to maintain and manipulate relevant information and to shield the system from distraction and interference during retention (Burgoyne & Engle, 2020; Cowan, 2008; Engle, 2018; Oberauer, 2009). Working memory is also essential for the maintenance and manipulation of abstract rules in both humans and non-human primates (Bunge, 2004; Mansouri et al., 2020; Nakahara et al., 2002). Such abstract rules underpin many higher-order cognitive abilities; indeed, unlike STM, working memory has been shown to be related to measures of general intelligence and executive functioning in humans (Conway et al., 2003; Engle, 2018; Miyake et al., 2000; Unsworth & Engle, 2007). As such, phylogenetic analysis of working memory in non-human primates would prove valuable to our understanding of the evolution of cognitive processes thought essential for goal-directed behavior. Some simple extensions to our current design to ensure the recruitment of working memory (cf. STM) processes could be to fill the retention interval of our memory paradigm with a distraction task, utilize a delayed match-to-sample task with distraction (e.g., Basile & Hampton, 2013), or utilize a version of the self-ordered search task (e.g., Petrides, 1995).

The results of our phylogenetic analysis are of course conditional on the sample we tested. Despite a concerted effort to obtain a large, diverse sample of primates, data collection for the present study was nevertheless influenced by the relative abundance of some species of primates over others in captivity, and our relative ability to test them. Certain taxa such as chimpanzees, capuchin monkeys, squirrel monkeys,

and long-tailed macaques were tested more frequently because they are common in research facilities and zoological institutions and are typically housed in large groups. Other taxa, for instance, pair-living gibbons or tamarins, had much smaller sample sizes (some as low as only one individual per species). To be more fully representative of the primate order, future research may have to expend additional efforts to recruit understudied species. Most importantly, to increase representativeness, future research should aim to also study primates living in the wild in addition to captive settings. Testing wild individuals and comparing their performance with the results presented here would allow us to evaluate the extent of a potential captivity effect on STM (Meulman et al., 2012). In addition, it is possible that species-specific social and ecological factors—which we found unrelated to variation in task performance in captivity—would have a stronger effect in wild primate STM. A transfer to non-captive settings, however, would come with significant methodological, logistical, and ethical challenges that would require collaboration beyond the ManyPrimates project.

Finally, the alignment between social and ecological predictor variables and phylogeny in our data limited the explanatory power of the modeling challenge. This alignment might have resulted partly from our opportunistic sampling strategy. Phylogenetic targeting (Arnold & Nunn, 2010; MacLean et al., 2012) of species before data collection in future studies will help to overcome this limitation, as this approach allows identification of species that would increase the power to detect correlations between the task performance and certain species-level characteristics.

One of the aims of the present study, and of the ManyPrimates project more generally, is to standardize testing methodologies across species and institutions to obtain comparable data from large and diverse samples. This approach allows us to assess how comparable results for a given species are at different sites. Figure S5 in the supplementary material shows the performance of six species for which we collected sufficient data at least 2 different sites. This comparison suggests that, even though there is variation across sites, species tend to perform on similar levels at different sites.

Despite our efforts to standardize data collection, there were some aspects of the testing protocol that could not be standardized across testing institutions. For example, subjects at some institutions were used to being separated from conspecifics for short periods of time and could be tested individually, whereas others had to be tested in their group or with their dependent offspring (see Supplementary Material). Furthermore, the group sizes of the tested individuals varied between and within species and across sites. Similarly, cup distance and board size could not be standardized across institutions due to the specific materials available at each site. However, we controlled for these variables in our analyses, and we did not find that substantive conclusions depend on the inclusion of these covariates. Another possible factor that could have influenced the subjects' task performance is the varying familiarity of the test subjects with the experimenters. However, given that this project was an intergroup-intersite collaboration, it was unavoidable and necessary that different experimenters collected the data at different locations.

The positive effect of task experience (i.e., previous participation in object choice tasks) on STM performance contrasts with the results of our pilot study, in which familiarity with object-choice tasks did not lead to a higher probability of success (ManyPrimates et al., 2019a). One potential explanation for such divergent results could be the more variable nature of the present dataset, both in terms of task experience and performance. Task experience can confer processing advantages by minimizing resource allocation to procedural details and thus maximizing resource allocation to relevant stimulus features. Recently, researchers have suggested that task experience may introduce a considerable bias in cross-species comparisons and may lead to unwarranted conclusions about the cognitive limitations of certain species, and thus, about cognitive evolution (see, e.g., Leavens et al., 2019). The importance of controlling for task experience is further underscored by the recent adoption of the so-called STRANGE (acronym for Social background; Trappability and self-selection; Rearing history; Acclimation and habituation; Natural changes in responsiveness; Genetic make-up; and Experience) framework within the field of ethological research (Rutz & Webster, 2021).

The positive linear relation between STM performance and relative age runs against our predictions and previous studies of STM in humans and primates. Previous studies report an inverted U-curve relationship between STM capacity and age, whereby STM capacity is highest in young adults and

decreases again with old age (Brockmole & Logie, 2013, Darusman et al., 2014). Such divergent results could be explained by different sampling strategies (i.e., opportunistic sampling in our study as opposed to targeted sampling in studies addressing age-related changes in STM). Since an age-related decline in STM capacity is typically reported for advanced ages (Brockmole & Logie, 2013), it is possible that our results reflect an underrepresentation of such older individuals in our sample. Whether or not this was the case is difficult to say because our longevity estimates (which we used to norm age) might be biased due to differences in life expectancies between wild and captive populations. For understudied primates, these estimates are also likely to be imprecise.

Conclusion

The present study used the largest and most diverse sample of primates ever included in an experimental study, a scientific achievement only made possible by the large-scale international collaboration among scientists and institutions that represents the ManyPrimates project. This dataset allowed us to conduct robust phylogenetic analyses on the evolution of primate short-term memory as well as to evaluate how well species-specific social and ecological factors predict performance in a delayed-response task. When aggregating the data across species, we found that primates – as a group – performed well above chance in the delayed response task. Task success was found to decrease as the delay between stimuli presentation and response increased and previous task experience led to a higher probability of success in the task. Phylogenetic relatedness was found to strongly predict the observed variation in task performance among species whereas none of the species-specific ecological or social variables considered substantially contributed in explaining such variation. We encourage future studies to build upon the dataset and analyses of the present study by investigating the effects that other species-level predictors as well as individual-level predictors (such as rearing history, previous experience in cognitive tasks or hierarchical rank) have on STM. We want to emphasize again that such comprehensive evolutionary analyses are not possible with smaller or less diverse datasets and therefore crucially depend on the large-scale collaborative approach taken in this project. We therefore hope that the work presented here can act as a model for future studies to investigate the evolution of other cognitive abilities in the primate lineage.

Author Contributions: *Conceptualization*: Manuel Bohn (MB)*; Investigation*: Géraud Aguenounon (GA), Matthias Allritz (MA), Sébastien Ballesta (SB), Alice Beaud (AB), Manuel Bohn (MB), Sally L. Bornbusch (SLB), Angela Brandão (AB), Thomas Bugnyar (TB), Léa Bustamante (LB), Josep Call (JC), Chuangshi Cao (CC), Kai R. Caspar (KRC), Sarah E. DeTroy (SED), Shona Duguid (SD), Timothy M. Eppley (TME), Claudia Fichtel (CF), Juan Felipe Reátiga Parrish (JFRP), Chi Gong (CG), Nicholas M. Grebe (NMG), R. Adriana Hernandez-Aguilar (RAHA), Daniel Hanus (DH), Daniel B.M. Haun (DBMH), Lou M. Haux (LMH), Annabella Helman (AH), Istvan Hernadi (IH), Esther Herrmann (EH), Lydia M Hopper (LMH), Lei Huang (LH), Sarah M. Huskisson (SMH), Ivo Jacobs (IJ), Zhiyong Jin (ZJ), Marine Joly (MJ), Fumihiro Kano (FK), Stefanie Keupp (SK), Evelin Kiefer (EK), Baláazs Knakker (BK), Katalin Kóczán (KK), Larissa Kraus (LK), Sze Chai Kwok (SCK), Marie Lefrançois (ML), Elizabeth V. Lonsdorf (EVL), Louise Loyant (LL), Siyi Liu (SL), Katarzyna Majecka (KM), Hélène Meunier (HM), Luca Morino (LM), Flávia Mobili (FM), Alba Motes-Rodrigo (AMR), Caroline Nkoy Ihomi (CNI), Dariusz Pietraszewski (DP), Anthony Roig (AR), Vedrana Šlipogor (VŠ), Alejandro Sánchez-Amaro (ASA), Yutaro Sato (YS), Allie E. Schrock (AES), Amanda Seed (AMS), Caroline L. Shearer (CLS), Diana da Silva (DS), Yanjie Su (YS), Jingzhi Tan (JT), Christoph J. Völter (CJV), Elizabeth Warren (EW)*; Data curation*: Manuel Bohn (MB), Kai R. Caspar (KRC), Julia Watzek (JW), Yseult Héjja-Brichard (YHB), Alba Motes-Rodrigo (AMR), Vincent Nijman (VN), Camille A. Troisi (CAT), Luke Maurits (LM), Drew Altschul (DMA), Kirsten Sutherland (KS)*; Methodology*: Manuel Bohn (MB), Christoph J. Völter (CJV)*; Formal analysis*: Manuel Bohn (MB), Luke Maurits (LM)*; Project administration*: Manuel Bohn (MB), Christoph J. Völter (CJV), Julia Watzek (JW), Drew Altschul (DMA); *Visualization*: Manuel Bohn (MB), James A. Grange (JAG), Julia Watzek (JW), Luke Maurits (LM)*; Writing – original draft*: Manuel Bohn (MB), Kai R. Caspar (KRC), Shona Duguid (SD), R. Adriana Hernandez-Aguilar (RAHA), Marine Joly

(MJ), Miquel Llorente (ML), Alba Motes-Rodrigo (AMR), Tomas Persson (TP), Alejandro Sánchez-Amaro (ASA), Gabriela-Alina Sauciuc (GS), Manon K. Schweinfurth (MKS), Derry Taylor (DT), Christoph J. Völter (CJV), Pauline Zablocki-Thomas (PZT)*; **Writing – review & editing**: Manuel Bohn (MB), James Brooks (JB), Charlotte Canteloup (CC), Kai R. Caspar (KRC), Alexandra Allison de Sousa (AAdS)**, Shona Duguid (SD), Timothy M. Eppley (TME), Claudia Fichtel (CF), Julia Fischer (JF), James A. Grange (JAG), Nicholas M. Grebe (NMG), Daniel B.M. Haun (DBMH), R. Adriana Hernandez-Aguilar (RAHA), Lydia M. Hopper (LMH), Ivo Jacobs (IJ), Marine Joly (MJ), Sze Chai Kwok (SCK), Miquel Llorente (ML), Julia Watzek (JW), Elizabeth V. Lonsdorf (EVL), Tomas Persson (TP), Alba Motes-Rodrigo (AMR), Manon K. Schweinfurth (MKS), Vedrana Šlipogor (VŠ), Alejandro Sánchez-Amaro (ASA), Gabriela-Alina Sauciuc (GS), Camille A. Troisi (CAT), Christoph J. Völter (CJV), Vincent Nijman (VN), Pauline Zablocki-Thomas (PZT), Laura Simone Lewis (LSL), Drew Altschul (DMA).

Conflict of Interest: Lydia M Hopper (LMH) is the Editor-in-Chief of *Animal Behavior and Cognition*, but this submission was handled independently by an Associate Editor and LMH was not involved in any aspect of the editorial process associated with this article (i.e. reviewer invitation and editor determinations and recommendations regarding reviews and acceptance). However, LMH did review the submission for formatting and copy editing after acceptance and prior to publication. Otherwise, the authors declare no competing interests.

References

Altschul, D., Terrace, H., & Weiss, A. (2016). Serial cognition and personality in macaques. *Animal Behavior and Cognition*, *3*(1), 46–64. https://doi.org/10.12966/abc.02.04.2016

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(3), 114–118. https://doi.org/10.1002/evan.20251

Arnold, C., & Nunn, C. L. (2010). Phylogenetic targeting of research effort in evolutionary biology. *American Naturalist*, *176*, 601–612. https://doi.org/10.1086/656490

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *225*(2), 82–90. https://doi.org/10.1038/scientificamerican0871-82

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*). *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(3), 239252. https://doi.org/10.1037/0097-7403.32.3.239

Bartus, R. T., Fleming, D., & Johnson, H. R. (1978). Aging in the rhesus monkey: Debilitating effects on short-term memory. *Journal of Gerontology*, *33*(6), 858–871. https://doi.org/10.1093/geronj/33.6.858

Basile, B. M., & Hampton, R. R. (2013). Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition*, *126*(3), 391–396. https://doi.org/10.1016/j.cognition.2012.10.012

Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology. General*, *144*(1), 85–102. https://doi.org/10.1037/xge0000031

Beran, M. J., Perdue, B. M., Futch, S. E., Smith, J. D., Evans, T. A., & Parrish, A. E. (2015). Go when you know: Chimpanzees' confidence movements reflect their responses in a computerized memory task. *Cognition*, *142*, 236–246. https://doi.org/10.1016/j.cognition.2015.05.023

Blomberg, S. P., Garland, T. J., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.

Boddy, A. M., McGowen, M. R., Sherwood, C. C., Grossman, L. I., Goodman, M., & Wildman, D. E. (2012). Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain scaling. *Journal of Evolutionary Biology*, *25*(5), 981–994. https://doi.org/10.1111/j.1420-9101.2012.02491.x

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: A study of 55,753 participants aged 8-75. *Frontiers in Psychology*, *4*, 12. https://doi.org/10.3389/fpsyg.2013.00012

Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 564–579. https://doi.org/10.3758/cabn.4.4.564

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(5), 175–186. https://doi.org/https://doi.org/10.1002/evan.20222

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630. https://doi.org/10.1177/0963721420969371

Byrne, R. W. (1997). The Technical Intelligence hypothesis: An additional evolutionary stimulus to intelligence? In A. White & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 289–311). Cambridge University Press.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans.* Clarendon Press/Oxford University Press.

Cacchione, T., & Rakoczy, H. (2017). Comparative metaphysics: Thinking about objects in space and time. In J. Call, J. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Perception, learning, and cognition* (pp. 579–599). American Psychological Association.

Constantinidis, C., & Procyk, E. (2004). The primate working memory networks. *Cognitive, Affective & Behavioral Neuroscience*, *4*(4), 444–465. https://doi.org/10.3758/cabn.4.4.444

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Cunningham, E., & Janson, C. (2007). A socioecological perspective on primate cognition, past and present. *Animal Cognition*, *10*(3), 273–281. https://doi.org/10.1007/s10071-007-0078-3

Darusman, H. S., Call, J., Sajuthi, D., Schapiro, S. J., 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

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not by sociality. *Nature Ecology & Evolution*, *1*(5), 0112. https://doi.org/10.1038/s41559-017-0112

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology*, *22*(3), 271–294. https://doi.org/10.1002/dev.420220307

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344–1347. https://doi.org/10.1126/science.1145463

Egelkamp, C. L., Jacobson, S. L., Cronin, K. A., Wagner, K. E., Ross, S. R., & Hopper, L. M. (2019). A comparison of sequential learning errors made by apes and monkeys reveals individual but not species differences in learning. *International Journal of Comparative Psychology*, *32*. https://escholarship.org/uc/item/024578f5

Elmore, L. C., & Wright, A. A. (2015). Monkey visual short-term memory directly compared to humans. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*(1), 32–38. https://doi.org/10.1037/xan0000050

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Geurten, M., Vincent, E., Van der Linden, M., Coyette, F., & Meulemans, T. (2016). Working memory assessment: Construct validity of the Brown-Peterson Test. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement*, *48*(4), 328–336. https://doi.org/10.1037/cbs0000057

Ghazizadeh, A., Griggs, W., Leopold, D. A., & Hikosaka, O. (2018). Temporal–prefrontal cortical network for discrimination of valuable objects in long-term memory. *Proceedings of the National Academy of Sciences of the USA*, *115*(9), 2135–2144. https://doi.org/10.1073/pnas.1707695115

Grimault, S., Robitaille, N., Grova, C., Lina, J.-M., Dubarry, A.-S., & Jolicoeur, P. (2009). Oscillatory activity in parietal and dorsolateral prefrontal cortex during retention in visual short-term memory: Additive effects of spatial attention and memory load. *Human Brain Mapping*, *30*(10), 3378–3392. https://doi.org/10.1002/hbm.20759

Harlow, H. F., & Bromer, J. A. (1939). Comparative behavior of primates. VIII. The capacity of Platyrrhine monkeys to solve delayed reaction tests. *Journal of Comparative Psychology*, *28*(2), 299–304. https://doi.org/10.1037/h0053770

Harlow, H. F., Uehling, H., & Maslow, A. H. (1932). Comparative behavior of primates. I. Delayed reaction tests on primates from the lemur to the orang-outan. *Journal of Comparative Psychology*, *13*(3), 313–343. https://doi.org/10.1037/h0073864

Herculano-Houzel, S., Collins, C. E., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences of the USA*, *104*(9), 3562–3567. https://doi.org/10.1073/PNAS.0611396104

Hopper, L. M., van de Waal, E., & Caldwell, C. A. (2018). Celebrating the continued importance of "Machiavellian Intelligence" 30 years on. *Journal of Comparative Psychology*, *132*(4), 427–431. https://doi.org/10.1037/com0000157

Hunter, W. S. (1913). The delayed reaction in animals and children. *Animal Behavior Monographs*, *2*, *1*, 86.

Inoue, S., & Matsuzawa, T. (2009). Acquisition and memory of sequence order in young and adult chimpanzees (*Pan troglodytes*). *Animal Cognition*, *12*, 59-69. https://doi.org/10.1007/s10071-009-0274-4

Jolly, A. (1966). Lemur social behaviour and primate intelligence. *Science*, *153*(3735), 501–506. https://doi.org/10.1126/science.153.3735.501

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Kappeler, P. (1999). Convergence and divergence in primate social systems. In J. Fleagle, C. Janson, & K. Reed (Eds.), *Primate communities* (pp. 158-170). Cambridge University Press. https://doi:10.1017/CBO9780511542381.010

Kawai, N., & Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature*, *403*(6765), 39–40. https://doi.org/10.1038/47405

Krupenye, C., & Call, J. (2019). Theory of mind in animals: Current and future directions. *WIREs Cognitive Science*, *10*(6), e1503. https://doi.org/https://doi.org/10.1002/wcs.1503

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2019). The mismeasure of ape social cognition. *Animal Cognition*, *22*(4), 487–504. https://doi.org/10.1007/s10071-017-1119-1

Lewis, A., Berntsen, D., & Call, J. (2019). Long-term memory of past events in great apes. *Current Directions in Psychological Science*, *28*(2), 117–123. https://doi.org/10.1177/0963721418812781

MacLean, E. L. (2016). Unraveling the evolution of uniquely human cognition. *Proceedings of the National Academy of Sciences of the USA*, *113*(23), 6348–6354. https://doi.org/10.1073/PNAS.1521270113

MacLean, E. L., Matthews, L. J., Hare, B. A., Nunn, C. L., Anderson, R. C., Aureli, F., Brannon, E. M., Call, J., Drea, C. M., Emery, N. J., Haun, D. B. M., Herrmann, E., Jacobs, L. F., Platt, M. L., Rosati, A. G., Sandel, A. A., Schroepfer, K. K., Seed, A. M., Tan, J., …Wobber, V. (2012). How does cognition evolve? Phylogenetic comparative psychology. *Animal Cognition*, *15*(2), 223–238. https://doi.org/10.1007/s10071-011-0448-8

Mansouri, F. A., Freedman, D. J., & Buckley, M. J. (2020). Emergence of abstract rules in the primate brain. *Nature Reviews Neuroscience*, *21*(11), 595–610. https://doi.org/10.1038/s41583-020-0364-5

ManyPrimates, Altschul, D., Bohn, M., Canteloup, C., Ebel, S. J., Hernández-Aguilar, R. A., Joly, M., Keupp, S., Petkov, C. I., Llorente, M., Proctor, D., Motes Rodrigo, A., Sutherland, K., Szabelska, A., Taylor, D., Völter, C. J., & Wiggenhauser, N. G. (in press). ManyPrimates: Collaboration and open science initiatives in primate research. In M. J. Beran & B. Schwartz (Eds.), *Primate Cognitive Studies*. Cambridge University Press. https://doi.org/10.31219/osf.io/7c93a

ManyPrimates, Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., …Watzek, J. (2019a). Establishing an infrastructure for collaboration in primate cognition research. *PLoS One*, *14*(10), e0223675. https://doi.org/10.1371/journal.pone.0223675

ManyPrimates, Altschul, D. M., Beran, M. J., Bohn, M., Caspar, K. R., Fichtel, C., Försterling, M., Grebe, N. M., Hernandez-Aguilar, R. A., Kwok, S. C., Llorente, M., Motes-Rodrigo, A., Proctor, D., Sánchez-Amaro, A., Simpson, E. A., Szabelska, A., Taylor, D., van der Mescht, J., Völter, C. J., & Watzek, J. (2019b). Collaborative open science as a way to reproducibility and new insights in primate cognition research. *Japan Psychological Research*, *62*, 205–220. https://doi.org/10.31234/osf.io/8w7zd

ManyPrimates, Géraud S. Aguenounon, S. B., Beaud, A., Bustamante, L., Canteloup, C., Joly, M., Loyant, L., Meunier, H., Roig, A., Troisi, C. A., & Zablocki-Thomas, P. (2020). ManyPrimates: Une infrastructure de

collaboration internationale dans la recherche en cognition des primates. *Revue de Primatologie*, *11*. https://doi.org/10.4000/primatologie.8808

Maslow, A. H., & Harlow, H. F. (1932). Comparative behavior of primates. II. Delayed reaction tests on primates at Bronx Park Zoo. *Journal of Comparative Psychology*, *14*(1), 97–107. https://doi.org/10.1037/h0072093

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Mercer, T., & McKeown, D. (2014). Decay uncovered in nonverbal short-term memory. *Psychonomic Bulletin & Review*, *21*, 128–135. https://doi.org/10.3758/s13423-013-0472-6

Meulman, E. J., Sanz, C. M., Visalberghi, E., & van Schaik, C. P. (2012). The role of terrestriality in promoting primate technology. *Evolutionary Anthropology*, 21(2), 58–68. https://doi.org/10.1002/evan.21304

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Milton, K. (1981). Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. *American Anthropologist*, *83*(3), 534–548. https://doi.org/10.1525/aa.1981.83.3.02a00020

Mittermeier, R. A., Wilson, D. E., & Rylands, A. B. (2013). *Handbook of the mammals of the world, Volume 3: Primates*. Lynx Edicions.

Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review*, *20*(1), 163–170. https://doi.org/10.3758/s13423-012-0313-z

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs*. R package (version 0.9.12-4.2). https://CRAN.R-project.org/package=BayesFactor.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756. https://doi.org/https://doi.org/10.1111/j.2041-210X.2012.00196.x

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*, *295*(5559), 1532–1536. https://doi.org/10.1126/science.1067653

Nissen, H. W., Riesen, A. H., & Nowlis, V. (1938). Delayed response and discrimination learning by chimpanzees. *Journal of Comparative Psychology*, *26*(2), 361–386. https://doi.org/10.1037/h0056387

Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Current Biology*, *19*(5), 190–191. https://doi.org/10.1016/j.cub.2009.01.010

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100. https://doi.org/10.1016/s0079-7421(09)51002-x

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Petrides, M. (1995). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *The Journal of Neuroscience*, *15*(1), 359–375. https://doi.org/10.1523/jneurosci.15-01-00359.1995

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences*, *1*, 40–46. https://doi.org/10.1016/j.cobeha.2014.08.004

Powell, L. E., Isler, K., & Barton, R. A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1865), 20171765. https://doi.org/10.1098/rspb.2017.1765

Prétôt, L., & Brosnan, S. F. (2019). Capuchin monkeys (*Cebus [Sapajus] apella*) show planning in a manual maze task. *Journal of Comparative Psychology*, *133*(1), 81–91. https://doi.org/10.1037/com0000133

R Core Team. (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Read, D. W. (2008). Working memory: A cognitive limit to non-human primate recursive thinking prior to Hominid evolution. *Evolutionary Psychology*, *6*(4), 676–71. https://doi.org/10.1177/147470490800600413

Resende, B., Ballesteros-Ardilla, A., Fragaszy, D., Visalberghi, E., & Izar, P. (2021). Revisiting the fourth dimension of tool use: How objects become tools for capuchin monkeys. *Evolutionary Human Sciences*, *3*, 18. https://doi.org/10.1017/ehs.2021.16

Riley, M. R., & Constantinidis, C. (2016). Role of prefrontal persistent activity in working memory. *Frontiers in Systems Neuroscience*, *9*, 181. https://doi.org/10.3389/fnsys.2015.00181

Rosati, A. G. (2017). Foraging cognition: Reviving the ecological intelligence hypothesis. *Trends in Cognitive Sciences*, *21*(9), 691–702. https://doi.org/10.1016/j.tics.2017.05.011

Rosati, A. G., & Santos, L. R. (2016). Spontaneous metacognition in rhesus monkeys. *Psychological Science*, *27*(9), 1181–1191. https://doi.org/10.1177/0956797616653737

Rutz, C., & Webster, M. M. (2021). Ethology adopts the STRANGE framework for animal behaviour research, to improve reporting standards. *Ethology*, *127*(2), 99–101. https://doi.org/10.1111/eth.13118

Sands, S. F., & Wright, A. A. (1980). Primate memory: Retention of serial list items by a rhesus monkey. *Science*, *209*(4459), 938–940. https://doi.org/10.1126/science.6773143

Sanz, C. M., Call, J., & Boesch, C. (2013). *Tool use in animals: Cognition and ecology*. Cambridge University Press. https://doi.org/10.1017/cbo9780511894800

Schwartz, B. L. (2019). Using natural ecology to predict higher cognition in human and non-human primates. *Animal Behavior and Cognition*, *6*(4), 344-354. https://doi.org/10.26451/abc.06.04.13.2019

Shettleworth, S. J. (2010). *Cognition, evolution and behavior*. Oxford University Press.

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior*. Johns Hopkins University Press.

Smuts, B. B., Cheney, D. L., Seyfarth, R. M., & Wrangham, R. W. (1986). *Primate societies*. The University of Chicago Press.

Sol, D. (2009). Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters*, *5*(1), 130–133. https://doi.org/10.1098/rsbl.2008.0621

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/10.1037/0033-295x.114.1.104

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In *The handbook of memory disorders* (pp. 249–270). Wiley.

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Perception, learning, and cognition.* (pp. 643–671). American Psychological Association. https://doi.org/10.1037/0000012-029

Wise, S. (2017). The evolution of the prefrontal cortex in early primates and anthropoids. In J. H. Kaas (Ed.), *Evolution of Nervous Systems* (pp. 387–422). Academic Press. https://doi.org/10.1016/B978-0-12-820584-6.00028-3

Wynn, T., Hernandez-Aguilar, R. A., Marchant, L. F., & Mcgrew, W. C. (2011). "An ape's view of the Oldowan" revisited. *Evolutionary Anthropology: Issues, News, and Reviews*, *20*(5), 181–197. https://doi.org/https://doi.org/10.1002/evan.20323

Zewald, J., & Jacobs, I. (2021). Object permanence. In *Encyclopedia of animal cognition and behavior*. Springer. https://doi.org/10.1007/978-3-319-47829-6_2103-1

Zola-Morgan, S., & Squire, L. R. (1986). Memory impairment in monkeys following lesions limited to the hippocampus. *Behavioral Neuroscience*, *100*(2), 155–160. https://doi.org/10.1037//0735-7044.100.2.155

Supplementary Material

Methods

Table S1

Site Specific Adjustments, Testing Arrangements and Reliability Coding

Site	Species (scientific name)	Cup distance (cm)	Board length (cm)	Test setting	Previous task experience	N	K[1]	Data from pilot study[3]
Ape Cognition & Conservation Initiative	Bonobo (*Pan paniscus*)	20	50.8	Individual	yes	5	0.96	yes
Apenheul, Apeldoorn	Northern White Cheeked Gibbon (*Nomascus leucogenys*)	17	40	Individual	no	1	1	no
Breeding Base of Beijing Zoo	François' Langur (*Trachypithecus francoisi*)	15	56	Individual	no	2	NA	no
Duisburg Zoo	Siamang (*Symphalangus syndactylus*)	17	40	Group	no	1	1	no
	King Colobus (*Colobus polykomos*)	17	40	Group	no	3	1	no
Bioparc Doué la Fontaine	Siamang (*Symphalangus syndactylus*)	17	40	Individual	yes	1	1	no
Duke Lemur Center	Black and White Ruffed Lemur (*Varecia variegata*)	20	60	Individual	no	1	1	no
	Black and White Ruffed Lemur (*Varecia variegata*)	20	60	Individual	yes	7		
	Blue Eyed Black Lemur (*Eulemur flavifrons*)	20	60	Individual	yes	10	1	no
	Coquerel's Sifaka (*Propithecus coquereli*)	20	60	Individual	yes	9	1	no
	Crowned Lemur (*Eulemur coronatus*)	20	60	Individual	yes	8	0.90	no
	Mongoose Lemur (*Eulemur mongoz*)	20	60	Individual	yes	8	1	no
	Red Ruffed Lemur (*Varecia rubra*)	20	60	Individual	yes	8	0.94	no
	Ring-tailed Lemur (*Lemur catta*)	20	60	Individual	yes	8	0.95	no
Primate Center of East China Normal University	Rhesus Macaque (*Macaca mulatta*)	30	107	Individual	no	4	NA	no

	Brown Capuchin Monkey (*Sapajus apella*)	11	60	Individual	yes	25	0.97	yes
Edinburgh Zoo	Chimpanzee (*Pan troglodytes*)	18.5	64	Group	no	12	0.98	yes
	Squirrel Monkey (*Saimiri sciureus*)	11	60	Individual	yes	18	1	yes
Franklin and Marshall College	Brown Capuchin Monkey (*Sapajus apella*)	16.5	39.37	Individual	no	18	1	no
German Primate Center	Ring-tailed Lemur (*Lemur catta*)	10	65	Individual	no	7	0.97	yes
	Black-and White Ruffed Lemur (*Varecia variegata*)	10	63	Individual	no	7		
	Long-tailed Macaque (*Macaca fascicularis*)	10	53	Individual	yes	17	0.99	yes
	Long-tailed Macaque (*Macaca fascicularis*)	10	60	Individual	yes	11		
Gibbon Conservation Center	Javan Gibbon (*Hylobates moloch*)	15	94	Individual	yes	3	1	no
Grastyán Translational Research Center	Rhesus Macaque (*Macaca mulatta*)	21	53	Individual	yes	17	0.89	no
Heidelberg Zoo	Western Lowland Gorilla (*Gorilla gorilla gorilla*)	20	75	Group/Individual	no	2	1	yes
Kristiansand Zoo	Black Faced Spider Monkey (*Ateles chamek*)	10	74.5	Group	no	1	1	yes
	Ring-tailed Lemur (*Lemur catta*)	10	74.5	Group/Individual	no	2		
Kumamoto Sanctuary	Bonobo (*Pan paniscus*)	15	60	Group/Individual	no	6	1	yes
	Chimpanzee (*Pan troglodytes*)	15	60	Group/Individual	no	6		
Lagos Zoo	Emperor Tamarin (*Saguinus imperator*)	10	35	Individual	no	1	0.89	no
	Golden Handed Tamarin (*Saguinus midas*)	10	35	Individual	no	1		
Language Research Center	Brown Capuchin Monkey (*Sapajus apella*)	10	56	Individual	yes	21	0.97	yes
Leipzig Zoo / Wolfgang Köhler Primate Research Center	Bonobo (*Pan paniscus*)	29	78	Group/Individual	yes	9	1	no
	Chimpanzee (*Pan troglodytes*)	29	78	Group/Individual	no	2	1	yes

	Species							
	Chimpanzee (*Pan troglodytes*)	29	78	Group/Individual	yes	20		
	Crowned Lemur (*Eulemur coronatus*)	19	59	Individual	no	3	1	no
	Diana Monkey (*Cercopithecus diana*)	19	59	Group/Individual	no	3	1	no
	Diana Monkey (*Cercopithecus diana*)	19	59	Group/Individual	yes	2		
	Golden Lion Tamarin (*Leontopithecus rosalia*)	19	59	Group/Individual	no	2	1	no
	Hamlyn's Monkey (*Cercopithecus hamlyni*)	19	59	Individual	no	2	1	no
	Hamlyn's Monkey (*Cercopithecus hamlyni*)	19	59	Individual	yes	1		
	Sumatran Orangutan (*Pongo abelii*)	29	78	Individual	yes	5	1	yes
	Western Lowland Gorilla (*Gorilla gorilla gorilla*)	30	78	Individual	no	1	0.78	no
	Western Lowland Gorilla (*Gorilla gorilla gorilla*)	30	78	Individual	yes	3		
Lincoln Park Zoo	Allen's Swamp Monkey (*Allenopithecus nigroviridis*)	11	30	Individual	no	2	0.78	no
	Western Lowland Gorilla (*Gorilla gorilla gorilla*)	19	61.5	Individual	no	3	0.9	yes
Łódź Zoo	Lion Tailed Macaque (*Macaca silenus*)	15	60	Group	no	1	1	no
	Lion Tailed Macaque (*Macaca silenus*)	15	60	Group	yes	3		
Monkey Haven, Isle of Wight	Rhesus Macaque (*Macaca mulatta*)	10	55	Group/Individual	yes	3	1	yes
	Barbary Macaque (*Macaca sylvanus*)	10	55	Individual	yes	3		
	East Javan Langur (*Trachypithecus auratus*)	10	55	Group	no	2	0.9	no
	Müller's Gibbon (*Hylobates muelleri*)	10	55	Group	no	1	1	no
	Siamang (*Symphalangus syndactylus*)	10	55	Group	no	1	1	no
	White Handed Gibbon (*Hylobates lar*)	10	55	Group	no	1	0.81	no

Paris Zoo	Brown Woolly Monkey (*Lagothrix lagotricha*)	10	70	Individual	no	3	1	no
	Black Lemur (*Eulemur macaco*)	25	60	Group	yes	1	1	no
	Brown Capuchin Monkey (*Sapajus apella*)	25	60	Group/Individual	yes	2	1	no
	Brown Lemur (*Eulemur fulvus*)	25	60	Group	yes	2	0.72	no
	Common Marmoset (*Callithrix jacchus*)	10	40	Individual	no	5	1	no
	Green Monkey (*Chlorocebus sabaeus*)	25	60	Group	no	4	0.89	no
Primate Center of Strasbourg University	Long-tailed Macaque (*Macaca fascicularis*)	25	60	Group	no	1	1	no
	Rhesus Macaque (*Macaca mulatta*)	25	60	Group	no	2	1	no
	Rhesus Macaque (*Macaca mulatta*)	25	60	Group	yes	2		
	Tonkean Macaque (*Macaca tonkeana*)	25	60	Group/Individual	yes	9	0.97	no
	White Faced Capuchin (*Cebus imitator*)	25	60	Group/Individual	no	7	0.86	no
	White Faced Capuchin (*Cebus imitator*)	25	60	Group/Individual	yes	1		
Lund University Primate Research Station Furuvik	Sumatran Orangutan (*Pongo abelii*)	20	60	Individual	yes	2	1	no
Primate Station, University of Zürich	Common Marmoset (*Callithrix jacchus*)	10	40	Individual	no	9	1	no
	Common Marmoset (*Callithrix jacchus*)	10	40	Individual	yes	7		
Texas Biomedical Research Institute	Olive Baboon (*Papio anubis*)	15	89	Group	no	6	NA	no
Animal Care Facility, University of Vienna	Common Marmoset (*Callithrix jacchus*)	15	52	Individual	no	2	1	no
	Common Marmoset (*Callithrix jacchus*)	15	52	Individual	yes	23		
Zoo Schönbrunn	Bornean orangutan (*Pongo pygmaeus*)	20	60	Individual	no	3	1	no
Wuhan Zoo	François' Langur (*Trachypithecus francoisi*)	19	60	Group	no	6	NA	no

Analysis

Prior Distributions for Parameters of the Confirmatory Analysis

Fixed effects

Intercept: `skewNormal(mu = 0, sd = 1.5, alpha = 3)`
It is unlikely that primates would perform below chance (p=1/3, logit ~ -0.69). Such values are less likely with this skewed prior.

Delay: `skewNormal(mu = -1.4, sd = 2, alpha = -3)`
It is unlikely based on prior work that longer delays lead to an increase in performance. Such positive effects of delay are less likely with this skewed prior.

Norm_age, Board_size, Cup_distance, Trial: `Normal(0,1)`
Informative prior because variables are continuous (most meaningful interpretable range is wider, from -3 to 3 for scaled predictors)

Task_experience: `Normal(0,2)`
Wider prior because predictor is categorical (interpretable range is from 0-1 only).

Standard deviation for random intercepts

Species, Site, Subject: `halfNormal(0,1)`
This prior allows for considerable variation in all random intercepts but makes extreme values, which are not meaningful in probability space, less likely.

Standard deviations for random slopes

Delay, Trial: `halfNormal(0,0.5)`

Correlation Matrices: `LKJ(1)`

Prior Distributions for Parameters of the Phylogenetic Analysis

For the phylogenetic analysis, we used the same priors. For species-level predictors, we used the following priors:

Continuous predictors (scaled): `N(0,1)`
Categorical: `N(0,2)`

Results

Confirmatory Analysis

Different Ways of Coding Delay

In the pre-registration, we noted that delay will be coded numerically, centered at 15s (medium) and with long delay as 1 and the short delay as -1. This way, only a single parameter is estimated to represent the effect of delay. It also assumes that the effect of delay is linear, that is, the difference between short and medium delay was the same as the difference between medium and long. We made this choice because it facilitated the inclusion of delay in the phylogenetic models as interactions with the additional predictor variables. When delay was coded as a three-level factor, additional interaction terms would have to be estimated for every level of the factor. Especially for the more complex models, this would have greatly increased the number of parameters in the model. In our pilot study, where we used a three-level coding, we found a significant difference between medium and long delay, which we also took as support for a numeric coding of delay.

When inspecting the data, we realized that the assumption of a linear relation between delay and performance did not hold in the current dataset. That is, the difference in performance between short and medium delay was much more pronounced compared to the difference between medium and long. The reason for this change compared to the pilot study is probably the much more diverse nature of the current sample, with a lot more species performing already close to chance in the short delay trials.

Nevertheless, the pre-registered model comparison with numeric delay favored the model including delay as a predictor. However, the difference in WAIC values compared to the standard errors of the WAIC value for each model was relatively small (see Table S2). As a consequence, we explored alternative ways of coding delay. On the one hand, as a three-level factor, on the other hand as a two-level factor with medium and long delay trials combined in one level. When comparing all three delay models to each other and a model without delay, we saw that factor-coding delay clearly improved the model, with an advantage for the three-level model.

For the confirmatory analysis, we therefore reported the results of the three-level factor model. For the phylogenetic analyses, however, we used the two-level coding because it avoids the issue with estimating additional interaction terms that the three-level coding brings. We would like to note that the results of the phylogenetic model comparison (i.e., the relative ordering of the different models) are the same with a numeric coding of delay.

Table S2

Model Comparison for Effects of the Different Delay Coding

Predictors	WAIC	SE	weight
delay (three-level factor)	14206.83	88.51	1.00
delay (two-level factor)	14229.28	87.98	0.00
delay (numeric)	14279.28	87.65	0.00
no delay	14788.29	80.55	0.00

Note. All models had the following basic structure: `correct ~ task_experience + cup_distance + board_size + trial +(1 + delay + trial | subject_site + site + species)`. The predictors in the table were added to this basic structure.

Additional Models for Age

As noted in the main text, a subset of the data came from a previously published study (ManyPrimates et al., 2019a). The confirmatory analysis in the previous paper differed from the present

one in that age was included as an interaction term with delay (assuming older individuals perform worse more with longer delays). Here, we compared this interaction model to the main effect model reported in the main text. In addition, we also add a model that assumes a quadratic relation between age and performance. It represents the hypothesis that performance increases with age but decreases again in old age. We coded delay as a two-level factor because we also included an interaction model.

Table S3 shows the result of the model comparison and suggests that the model positing a linear relation between age and performance makes the best out of sample predictions. Figure S1 visualizes the relation between age and short-term memory (hereafter STM) performance based on the data.

Table S3

Model Comparison for Different Age Effects

Predictors	WAIC	SE	weight
`delay + age`	14229.68	87.97	0.57
`delay + age + age`^{2}	14230.31	87.97	0.33
`delay * age`	14232.79	88.01	0.10

Note. All models had the following basic structure: `correct ~ task_experience + cup_distance + board_size + trial + (1 + delay + trial | subject_site + site + species)`. The predictors in the table were added to this basic structure.

Figure S1

Relation Between Age (Normed by the Maximum Recorded Life Span of the Species) and STM Performance (i.e., Proportion of Correct Responses in Medium and Long Delay Trials)
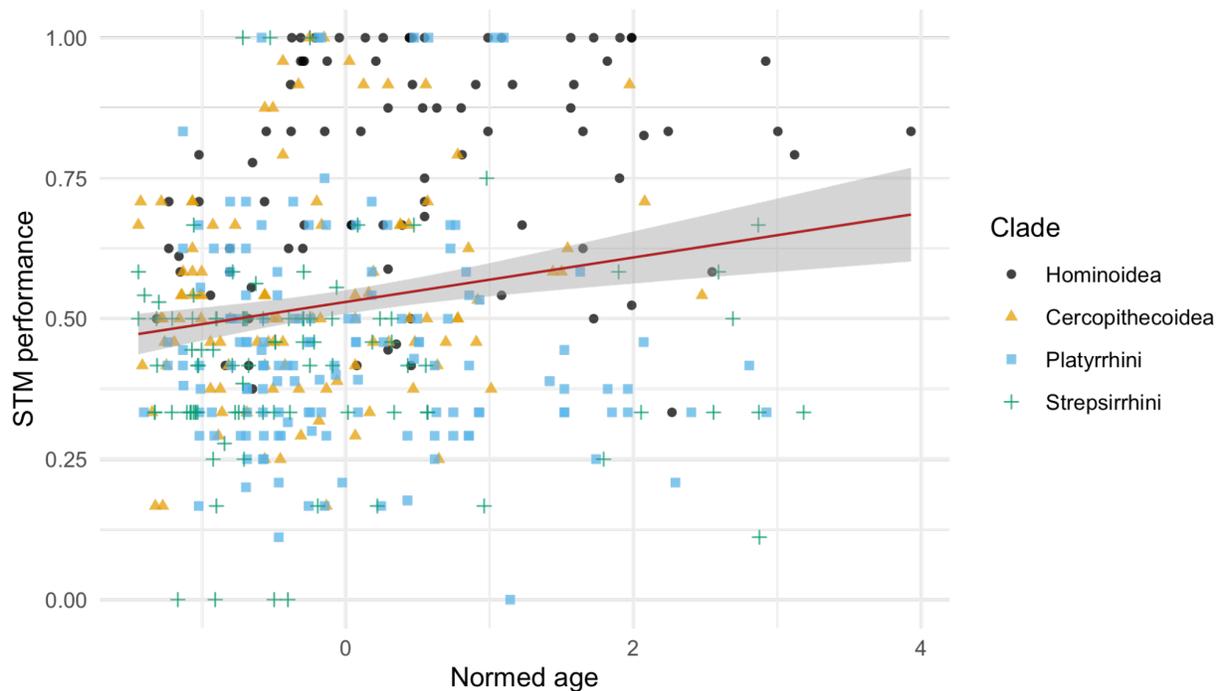

Note. The regression line (with 95% confidence interval in grey) is based on a linear model.

Relation Between Body Size and Cup Distance

In the main text, we suggested that the effect of cup distance in the confirmatory analysis might be driven by the fact that cup distance was correlated with body size. Figure S2 below visualizes this relation and shows a substantial correlation between the two variables. Body size, in turn, varied along phylogeny and so did performance (see main text and Table S4). We therefore suspect that cup distance does not have a direct causal effect on performance. Following a reviewer's suggestion, we re-ran the confirmatory model that included the ratio between cup distance and body size (cup distance/body size) as a predictor instead of cup distance. Larger values in this ratio reflect larger cup distance relative to body size. Thus, it (partially) accounts for the increase in cup distance due to body size. As expected, this ratio is largely unrelated to performance (β = -0.12, 95% CrI: -0.30 - 0.06).

Figure S2

Relation Between Body Size (Female Body Size, Same Value for all Individuals from a Species) and Cup Distance (Individual Specific)
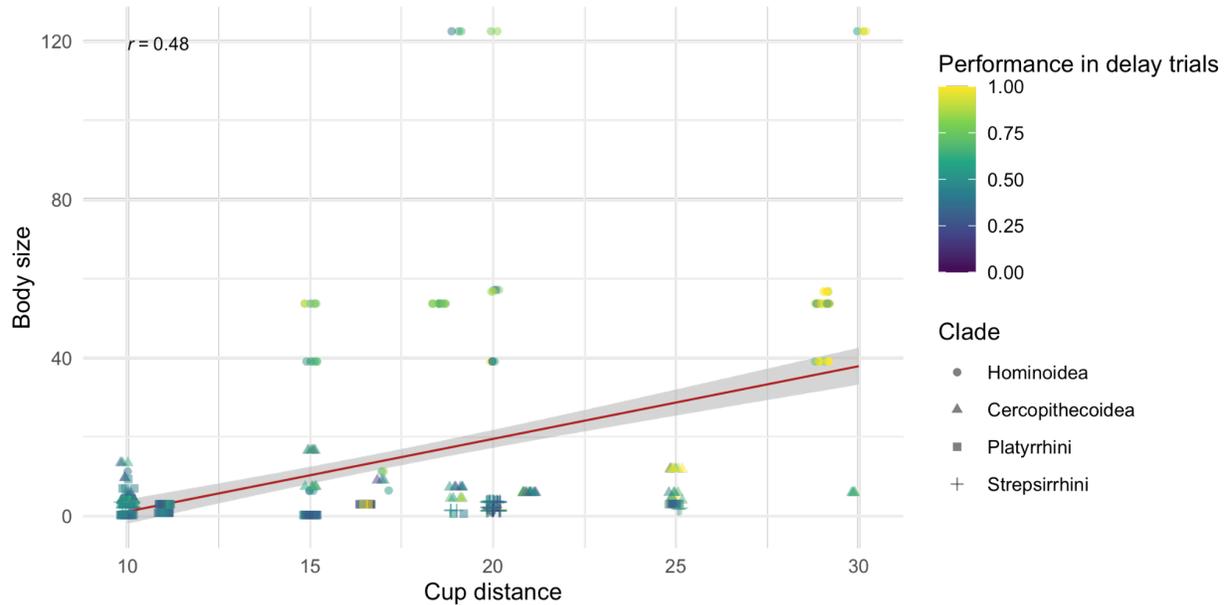
Note. Points are colored according to the average proportion of correct responses for each individual in the medium and long delay trials.

Phylogenetic Analysis

Prior Sensitivity Analysis for Phylogenetic Models

Our WAIC-based comparison of phylogenetic models attempts to answer the question of whether certain predictor variables can explain variation in STM performance over and above phylogeny. Any answer to this question necessarily depends on how much phylogeny is permitted to influence the model's predictions. In models which posit only minor differences in parameter values across different species there may be more variation in performance data "left over", which can be attributed to predictor variables. Since our models include prior distributions on the variance of all random effects of species, it is possible that different choices of these priors may change whether or not any models are assessed as outperforming the baseline. Therefore, we tried alternative comparisons of some models against the baseline using "tighter" priors on the variance of random effects of species, i.e., biasing the model away from strong effects of phylogeny. We compared the base, dietary breadth, home range and vocal repertoire models (i.e., the four

best performing models based on the results of the phylogenetic model comparison) against each other with the prior on all per-species random effects changed from N(0, 1.0) (as per the main comparison) to N(0, 0.5) and N(0, 0.25). The resulting WAIC comparisons did not differ in terms of the conclusions warranted: with all three choices of prior distribution, the differences in WAIC values between these four models were extremely small compared to the standard errors of the values, with parsimony therefore favoring the baseline model.

Exploratory Main Effects Models

Some of the models that were submitted to the modeling challenge (see main text) included interaction terms. In an exploratory phylogenetic model comparison, we disassembled these interaction models into separate main effects models (one model per predictor) and compared them to the other submitted models as well as to the baseline model. Our reasoning was that the added complexity of the interaction terms might have overshadowed the explanatory value of single predictors. Table S4 shows that including these models, however, did not change the results of the model comparison. The baseline model was ranked higher than all the models including a predictor.

Table S4

Phylogenetic Model Comparison Including Additional Fixed Effects Models

Model (predictors)	WAIC	SE	weight
`(baseline)model`	14091.57	87.58	0.14
`vocal_repertoire`	14091.70	87.66	0.13
`dietary_breadth`	14092.19	87.67	0.10
`group_size`	14092.83	87.68	0.07
`terrestriality`	14092.85	87.65	0.07
`home_range/body_size + feeding_budget * dietary_breadth`	14093.14	87.81	0.06
`feeding_budget`	14093.19	87.66	0.06
`percent_frugivory`	14093.21	87.67	0.06
`day_journey`	14093.36	87.66	0.06
`diurnal_resting`	14093.39	87.67	0.06
`home_range`	14093.40	87.70	0.06
`percent_frugivory * terrestriality`	14093.83	87.79	0.05
`color_vision`	14094.19	87.72	0.04
`day_journey_length * group_size`	14095.12	87.85	0.02
`diet_diversity`	14096.46	87.93	0.01

Note: All models had the same basic structure (baseline model): `correct ~ delay + (1 + delay + trial | subject_site + site + species)` with species capturing phylogeny. The predictors in the table were added to this baseline model as interactions with delay.

## *Model Parameters for Phylogenetic Models*

As mentioned in the main text and above, the baseline, vocal repertoire and dietary breadth models were indistinguishable from one another based on WAIC scores and weights. Furthermore, as shown in the prior sensitivity analysis above, the predictor models outperformed the baseline model when the effect of phylogeny was reduced. Below, we therefore report the relation between species vocal repertoire and

species dietary breadth and the performance in the delayed response task. Figure S3 visualizes the model parameters for the two predictor models and the baseline model.

In all models, the estimate for delay was negative, showing that longer delays led to worse performance (baseline: β = -0.48; ± 95% CrI: -0.76, -0.20; vocal repertoire: β = -0.46; ± 95% CrI: -0.68, -0.23; dietary breadth: β =-0.47; ± 95% CrI: -0.74, -0.19). The two predictors had a positive impact on the direction that a larger vocal repertoire or a broader diet were associated with better STM performance (vocal repertoire: β = 0.23; ± 95% CrI: -0.01, 0.47; dietary breadth: β = 0.31; ± 95% CrI: 0.14, 0.48). In the case of vocal repertoire, there was an additional negative interaction between delay and the predictor, suggesting that the positive effect of the predictor was weaker for longer delays (β = -0.10; ± 95% CrI: -0.18, -0.01). However, please note again that despite the fact that the posterior distributions for the key parameters suggested that there is a substantial relation between the predictor and STM abilities, including these predictors did *not* yield substantially better predictions compared to the baseline model.

Figure S3

Posterior Distribution for Model Predictors for the Baseline, Vocal Repertoire, and Dietary Breadth Models
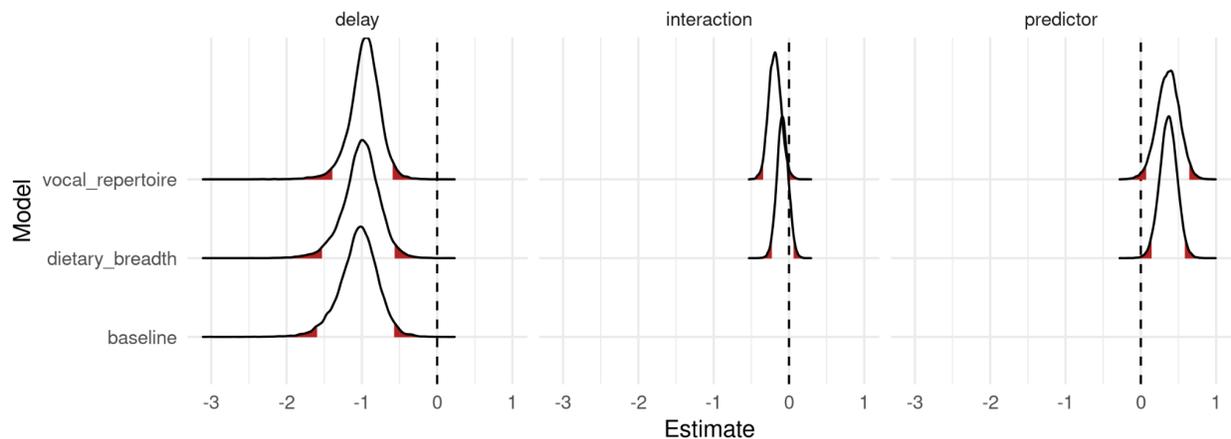
Note. Delay is the estimate for the main effect of delay, *interaction* gives the estimate for the interaction between delay and the predictor, and *predictor* is the main effect of the predictor variable. Red areas mark the 2.5% and 97.5% tails of each distribution.

Phylogenetic Signal in Predictor Variables

As we mentioned in the main text, one reason that the baseline model outperformed all predictor models might be the strong phylogenetic signal in the predictor variables themselves. That is, the predictor variables do not provide additional information about species' STM abilities once phylogeny has been accounted for. Table S5 reports the phylogenetic signal for the numerical predictor variables.

Table S5

*Phylogenetic Signal in Predictor Variables*

Predictor	Pagel's λ	Blomberg's K
vocal_repertoire	1.00	0.39
group_size	1.00	0.45
home_range	0.00	0.10
day_journey_length	0.18	0.32
resting_time_percent_activitybudget	0.82	0.37
feeding_budget	0.64	0.29
dietary_breadth	0.00	0.20
percent_frugivory	0.37	0.27
body_size	1.00	0.61

Correlations between Predictors and STM Abilities

Some of the predictor variables were substantially correlated with performance in the delayed-response task. Figure S4 visualizes these correlations for the numerical predictors.

Figure S4

Correlations between STM Performance and the Numerical Predictor Variables (Scaled and Centered)
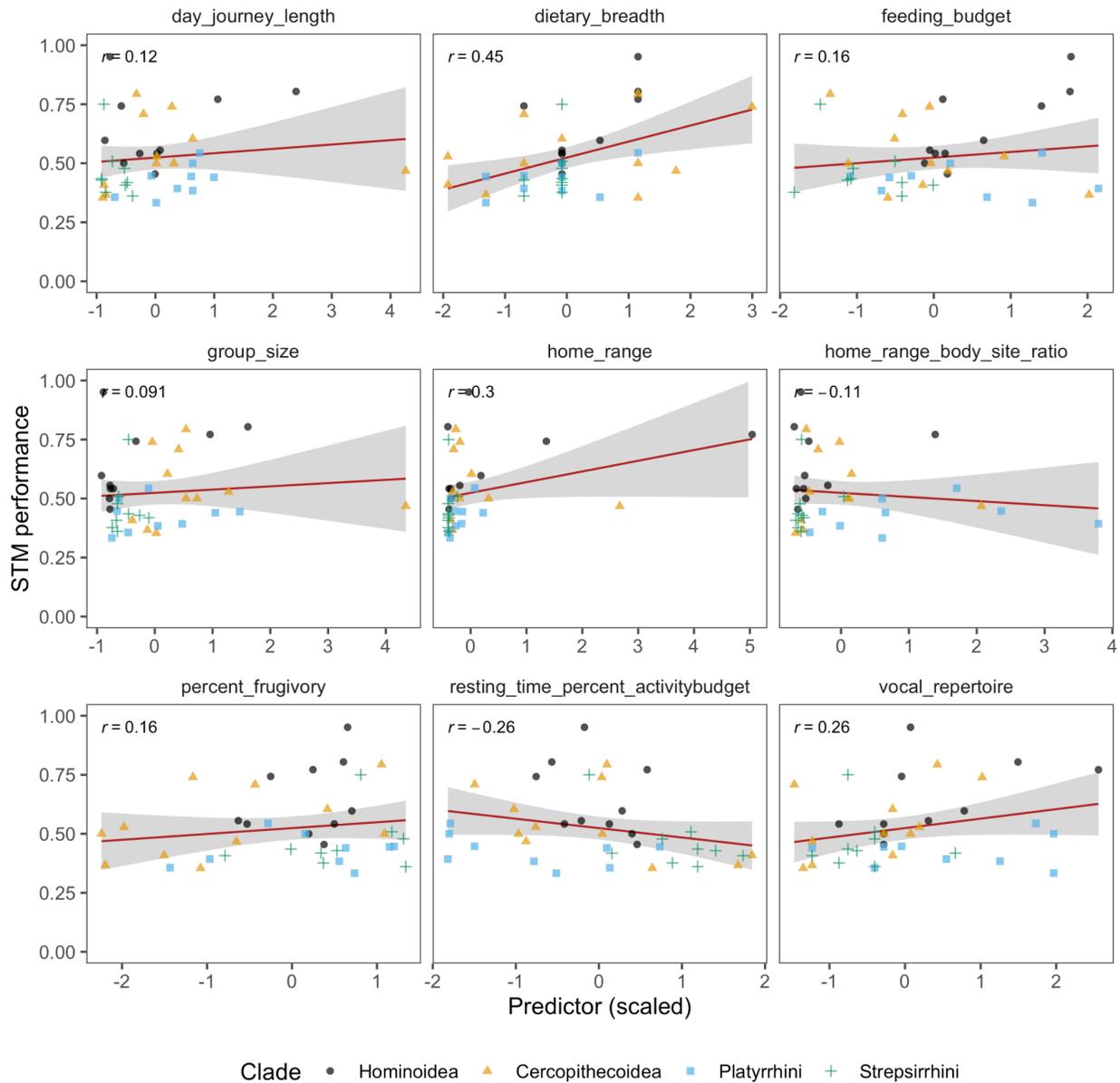
Note. STM abilities were computed by selecting the medium and long delay trials and averaging across them for each species. Coefficients are Pearson correlations.

Comparison Across Sites

Below we compare the performance of individuals from one species across the different data collection sites. This descriptive analysis gives an impression of how stable performance is. We only selected species for which there were at least five individuals tested per site.

For all species, performance was similar at different sites (Figure S5). Exceptions were the performance of bonobos in the long and medium delay condition and the performance of brown capuchin

monkeys in the long delay condition. For black-and-white ruffed lemurs there seemed to be a more systematic site effect.

Figure S5

Performance in the Delayed Response Task by Site for Species with More than One Data Collection Site and More than Five Individuals Tested per Site
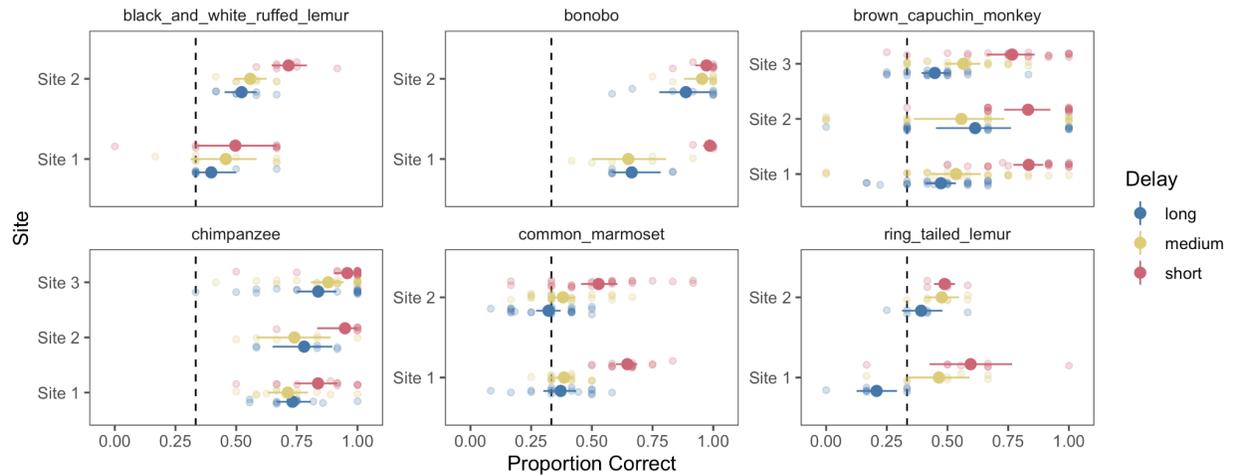
Note. Light points show individual means, solid points show group means with 95% confidence intervals.

Task Reliability

In addition to studying group-level variability, the delayed response task could be used to measure individual differences. To examine how well suited the task is for this purpose, we assessed the split-half reliability of the task. That is, for each individual, we split the data into odd and even trials, computed the mean performance for each of these test halves, and correlated them. Figure S6 visualizes the result and suggests an acceptable level of reliability. Of course, this result is subject to change if computed by species. For example, given that many chimpanzees performed at ceiling, there was much less variation to begin with. As a consequence, the split-half reliability would be much lower. We advise researchers to assess the reliability of the task in their sample before using it to study individual differences.

Figure S6

Split-Half Reliability for the Delayed Response Task
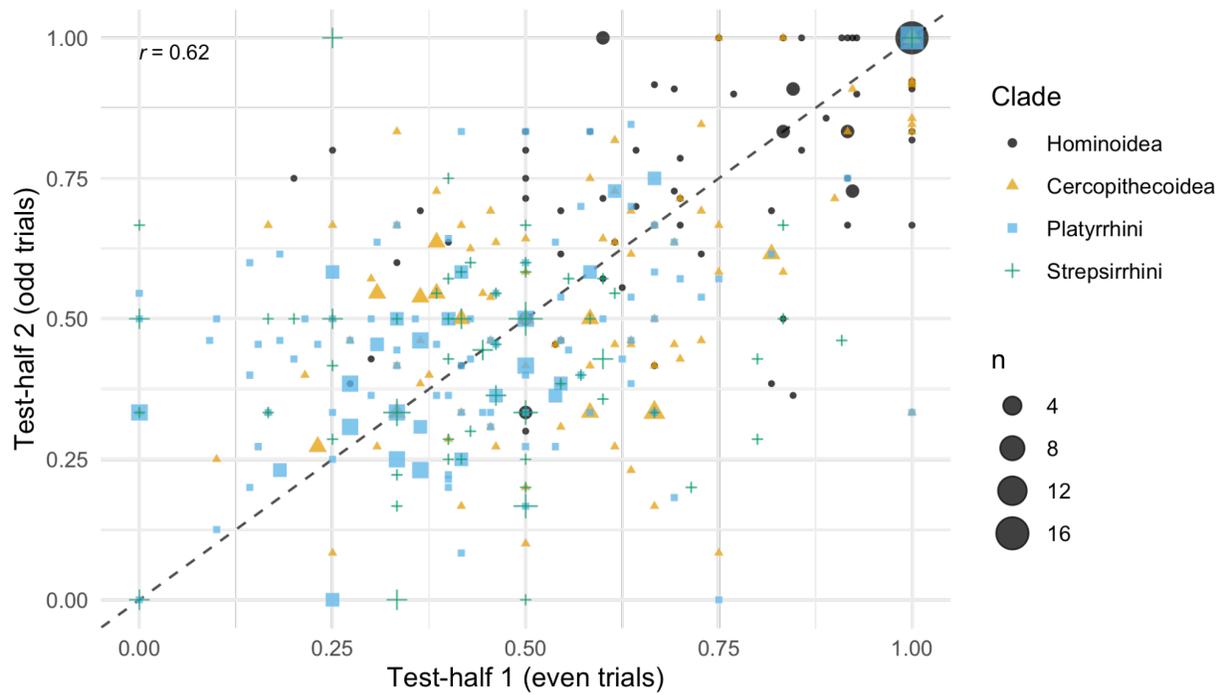
Note. Test halves were constructed by splitting the data into odd and even-numbered trials for each individual and computing the mean. The coefficient gives the Pearson correlation between the two test halves. Dashed line is the identity line.

Site Descriptions

Animal Care Facility, University of Vienna

Background

Marmoset Laboratory is a part of the Animal Care Facility, Department of Behavioral and Cognitive Biology, Faculty of Life Sciences, University of Vienna and is located in Biocenter, UZA I, Althanstrasse 14, Vienna, Austria. The laboratory consists of two animal keeping rooms and a large experimental room. The monkeys have access to all rooms via an interconnecting hallway with a tunnel system with moveable doors. Both keeping rooms usually house two family groups that are visually separated but remain in acoustic and olfactory contact. The socially unstable family groups are sometimes subdivided by a wire mesh into two smaller units for a limited amount of time, during which the sub-units remain in visual contact. Indoor temperature is kept between 21-29°C and humidity levels between 30-60%.

Animals

Twenty-five common marmosets (*Callithrix jacchus*) participated in testing, 9 females and 16 males, belonging to 5 family groups: Pooh (N = 2), Sparrow (N = 4), V group (N = 6), Veli (N = 8), and Kiri (N = 5). All animals lived in their social groups. No animal was involved in any invasive research.

Enclosures

The dimensions of the family group indoor-outdoor enclosures were approximately 5 m X 2.5 m X 2.5 m. The indoor enclosures had coniferous pellet bedding and both indoor and outdoor enclosures were equipped with tree branches, several sleeping structures (e.g., hammocks, hanging tunnels, baskets), other structures for climbing, swaying, playing, resting, gnawing or husbandry (e.g., wood boards, tires, cloth pieces, transport boxes, ropes), as well as additional enrichment objects that were regularly changed. An infrared lamp is attached to every indoor enclosure, to improve the well-being of animals. The animals had access to the outdoor enclosures during warm periods of the year, when the temperatures were above approximately 5°C. During the testing, habituation, or as enrichment, the animals had access to the small experimental enclosures within the laboratories and to the enclosures in the larger experimental room.

Diet

Subjects were never food or water deprived. In the morning before the testing sessions, the marmosets received monkey pellets. For older individuals the pellets were previously soaked in water. After the testing, i.e., around midday, the marmosets received their full lunch which was a mix of fruit, vegetables, marmoset jelly and gum, mealworms, eggs, cheese, or yogurt. As a special treat, the monkeys got crickets, granola, hanging fruit or foraging boxes with mealworms. As an incentive for participating in the study, subjects received small pieces of banana as a reward. Water was always available *ad libitum*, both in the home cages, as well as in the experimental cages during testing.

Ethical Approval

We obtained ethical approval from the Animal Ethics and Experimentation Board of Faculty of Life Sciences, license number 2020-015.

Research Permission

The keeping conditions for behavioral research were approved by the Austrian Federal Ministry of Science, Research and Economy (BMWFW), Geschäftszahl (GZ) BMWFW-66.006/0011-WF/II/3b/2014 from 22.05.2014; as our experiments were appetitive, non-invasive and based exclusively on behavioral tests, they were not classified as animal experiments under the Austrian Animal Experiments Act (§2. Federal Law Gazette No.501/1989).

Research Training

The research was carried out by two trained students from the University of Vienna in a close collaboration with lead researchers who had extensive experience with common marmosets, passed courses on common marmosets (i.e., EUPRIM-Net course on Marmosets as Animal Models) and with an accreditation in designing and performing cognitive and behavioral tests with primates (i.e., Laboratory Animal Science Course on Primates according to FELASA guidelines, Functions A & B, organized by the European Primate Network (EUPRIM-Net) at the German Primate Centre, Göttingen, Germany, under the Directive 2010/63/EU).

Research Participation

The animals participated on a voluntary basis in the tests. In particular, they entered the tunnel system and the small experimental cages voluntarily. If the animals showed signs of stress, the experiment would stop and continue on the next testing day.

Ape Cognition and Conservation Initiative

Background

The Ape Cognition and Conservation Initiative was formed in 2013 under the direction of Dr. William Hopkins and Dr. Jared Taglialatela. The facility is located on 230 acres of land outside of Des Moines, Iowa and currently houses five bonobos (*Pan paniscus*). Researchers from across the globe collaborate with ACCI staff and scientists to conduct research on great ape communication and cognition.

Animals

ACCI contributed data from five bonobos. Four males: Kanzi (38), Maisha (18), Teco (8), and Nyota (20). One female: Elikya (21). All five bonobos have lived at the facility since 2005.

Enclosures

The bonobos at ACCI are housed in 13 different indoor enclosures ranging in size from roughly 47 to 125 m^2. All enclosures are equipped with environmental enrichment and allow for both research and enrichment apparatuses to be added and removed. Bonobos can willingly separate into any of these spaces to participate in cognitive research or socialize in groups of two or more for behavioral research. There are two outdoor yards with 2.4 hectares of ape space where researchers can observe the apes from a birds-eye view. ACCI has roughly 262 m^2 of human-only areas, including office space, a public lobby, a kitchen area for preparing food for the apes, and a vet suite.

Diet

Animals are maintained on a veterinarian approved diet consisting of various fruits, vegetables, seeds, and nuts provided throughout each day in meals, foraging enrichment, and as rewards during cognitive testing. The bonobos are never water restricted.

Ethical Approval

The research was approved by the IACUC committee of ACCI (Protocol #170904-01R), no permit number issued. ACCI is certified by the Association of Zoos and Aquariums.

Research Permission

All research performed at ACCI is approved by ACCI's IACUC committee. Researchers must be listed on an approved IACUC document.

Research Training

All researchers and staff at ACCI have completed rigorous online and in-person training to safely work around apes. All visiting researchers have completed online and in-person training and are required to remain at least 1 m away from animal enclosures during testing.

Research Participation

All individuals participate in cognitive testing by voluntarily entering testing spaces and willingly separating themselves from other individuals, when necessary.

Other

There is no breeding program at place at ACCI. In the future, ACCI plans to introduce and house additional bonobos from various AZA-accredited facilities across the United States. After introduction into stable social groups, ACCI plans to include these individuals in future research programs.

Primate Park Apenheul, Apeldoorn

Background

Apenheul is a zoological garden in Apeldoorn, the Netherlands, specialized on keeping primates. It is internationally renowned for its primate husbandry and displays 35 species of lemurs and anthropoids. Apenheul is coordinating the European association of zoos and aquariums (EAZA) ex-situ program (EEP), a population management program, for the Western lowland gorilla (*Gorilla gorilla*), emperor tamarin (*Saguinus imperator*), woolly monkey (*Lagothrix sp.*) and Javan langur (*Trachypithecus auratus*).

Ethical approval

Thomas Bionda, MSc, Zoological manager.

Animals

Yuxi, a single male individual of the species *Nomascus leucogenys* (Northern white-cheeked gibbon) was tested. The subject was 8 years old and was kept in temporary isolation due to tensions experienced in his natal group, which is also housed at Apenheul. It was planned to rehome the subject to another zoological garden in coordination with the Northern white-cheeked gibbon EEP.

Enclosures

The outside enclosure measured 50 m^2 and was 5 m high. It connected to the inside enclosure via a mesh tunnel. A photograph of the outside enclosure, where testing took place, is presented below. Keepers regularly present the gibbon with behavioral enrichment such as food puzzles.

Diet

Subjects were never food or water deprived. The subject was kept on a diet of diverse vegetables, fruits and pellets, carefully selected by the nutritionist. As an incentive for participating in the study, apple slices were used. Apples constituted a regular part of its diet but proved to be highly desirable to the subject.

Research Permission

Research at Apenheul needs to be approved by the zoological manager and curator and has to comply with the standards of EAZA and the Nederlandse Vereniging van Dierentuinen (NVD).

Research Training

Responsible keepers give a basic introduction on interaction with and peculiarities of the respective subjects. Formal research training was not provided.

Research Participation

Animals must voluntarily approach the experimental setting in order to participate in scientific studies at Apenheul and are not forced to any extent to do so. Research was carried out at their regular enclosures, which ensured that subjects could recede from the experiments when desired.

Bioparc Doué la Fontaine

Background

Bioparc Doué la Fontaine is a zoological garden in Western France. Primates comprise an important part of the zoo's collection, especially gibbons and spider monkeys and the zoo is coordinating the EAZA EEP for the Colombian spider monkey (*Ateles fusciceps*) and the variegated spider monkey (*A. hybridus*).

Animals

A single subject, a juvenile male siamang (*Symphalangus syndactylus*, named Django) was tested at Bioparc Doué la Fontaine. Attempts to collect responses to the experimental set-up from his parents failed.

Enclosures

Siamangs were kept on an island enclosure with access to a heated indoor compound and were tested exclusively indoors.

Diet

Subjects were never food or water deprived. Siamangs were kept on a species-specific diet, including diverse vegetables and fruits. As an incentive for participating in the study, raisins were used, which proved to be highly desirable for the subject.

Ethical Approval

Dr. Florine Wedlarski (veterinarian / curator).

Research Permission

Research at Bioparc de Doué la Fontaine needs to be approved by the responsible board of curators as well as by the head keepers and has to comply with the standards of EAZA.

Research Training

Responsible keepers and veterinarians give a basic introduction on interaction with and peculiarities of the respective animals. Formal research training was not provided.

Research Participation

Animals must voluntarily approach the experimental setting in order to participate in scientific studies and are not forced to any extent to do so. Research was carried out at the regular enclosures, which ensures that subjects could recede from the experimenter if desired.

Breeding Base of Beijing Zoo

Background

Breeding Base of Beijing Zoo is located in Beijing, China. It is the breeding base of the oldest zoo in China (1906-present).

Animals

Three Francois' langurs (*Trachypithecus francoisi*) (1 male, 2 females) and two black snub-nosed monkeys (*Rinopithecus bieti*) were tested, but one Francois' langurs and two black snub-nosed monkeys completed < 9 trials and thus were not included in the study. Invasive research has never been conducted at the breeding base.

Enclosures

The animals have access to indoor enclosures (3 m X 4 m X 2 m) and outdoor enclosures (3 m X 4 m X 2 m), but the outdoor enclosures are not used in winter. They live in a social group of three subjects, with free access between three indoor enclosures. The outdoor enclosures are not kept open to the indoor enclosures.

Diet

Subjects were never food or water deprived. Daily diet included carrots, bananas, cucumbers and leaves. As an incentive for participating in the study, carrots and bananas were used. Because the quantity of food the subjects ate could not be increased, the experiments were run before they were fed.

Ethical Approval

The Breeding Base of Beijing Zoo gave permission for this study.

Research Permission

The research was approved by the zoo management. The study was also approved by the Ethics and Physical Protection Committee of the School of Psychological and Cognitive Sciences at Peking University.

Research Training

No systematic training was conducted. The animals were only trained (via positive reinforcement) to come when their names were called.

Research Participation

Testing happened in the indoor enclosures. The door to the outdoor enclosure was not open because Francois's langurs cannot adapt to the low winter temperature of Beijing.

Duisburg Zoo

Background

Duisburg Zoo is a zoological garden in Western Germany that is especially renowned for its marsupial and cetacean husbandry. However, primates traditionally constitute an important part of the zoo's collection with most species being kept at the Äquatorium building. Regarding primates, Duisburg Zoo is primarily focused on African ape and monkey species. The zoo is coordinating the EAZA EEP for the king colobus (*Colobus polykomos*).

Animals

Four species were studied at Duisburg Zoo. X group-housed king colobus monkeys (*Colobus polykomos*) of mixed sex was tested alongside a pair-housed white-cheeked gibbon female (*Nomascus leucogenys*, Sophie), a single-housed male siamang (*Symphalangus syndactylus*, Jupp) and three group-housed male white-naped mangabeys (*Cercocebus atys lunulatus*).

Enclosures

All species tested were kept at the Äquatorium building and could freely move between inside and outside enclosures, both enriched with numerous climbing structures as well as toys and food puzzles which were regularly changed by the keepers.

Diet

Subjects were never food or water deprived. They were kept on species-specific diets. In the case of the mangabeys, siamang and white-cheeked gibbons, these included diverse vegetables and fruits. King colobus were provided with fresh leaves collected by the keepers as well as pellets suitable for their specialized folivorous diet. Occasionally, vegetables were fed as dietary supplements. As an incentive for participating in the study, grapes (gibbons/mangabeys) and leaf-eater pellets (king colobus) were used. Both constitute a regular part of the diet of the respective subjects and proved to be highly desirable to the primates.

Ethical Approval

Dr. Carolin Bunert (veterinarian / curator), Alexander Nolte (head keeper)

Research permission

Research at Duisburg Zoo needs to be approved by the responsible head keeper and curator and has to comply with the standards of the EAZA.

Research Training

Responsible keepers give a basic introduction on interaction with and peculiarities of the respective subjects. Formal research training is not provided.

Research Participation

Animals must voluntarily approach the experimental setting in order to participate in scientific studies at Apenheul and are not forced to any extent to do so. Research is carried out at their regular enclosures, which ensures that subjects can recede from the experiments when desired.

Duke Lemur Center

Background

The Duke Lemur Center (DLC) was founded in 1966. With more than 200 animals across 14 species, the DLC houses the world's largest and most diverse population of lemurs outside their native Madagascar.

Animals

A total of 59 individuals across seven different species (*Propithecus coquereli, Eulemur flavifrons, Eulemur coronatus, Eulemur mongoz, Lemur catta, Varecia variegata,* and *Varecia rubra*) participated in the study. The DLC does not conduct invasive research with their animals.

Enclosures

DLC animals were housed socially, generally in large indoor/outdoor enclosures (23.2 –951.3 m^2, depending on group size), and were exposed to natural daylight and the local photoperiod. During the warmer months, some of the animals had access to larger, forested enclosures (0.6 –11 hectares), often with several species occupying the same habitat. Animals were temporarily separated from group-mates while participating in the trials.

Diet

Subjects were never food or water deprived. *Eulemur mongoz* and *Propichecus coquereli* were fed folivore chow, whereas the other *Eulemur* species and *Lemur catta* were fed monkey chow (Monkey Diet, LabDiet, St. Louis, MO, USA). All animals received fruits and vegetables to supplement their diet. The animals that range semi-free included local vegetation and insects they gathered from the forest. Food rewards were adjusted to be diet-appropriate for each species participating in the trials.

Ethical Approval

The Duke University Medical Center Institutional Animal Care and Use Committee (IACUC) gave ethical approval for the MP1 study (Protocol Registry Number A218-19-10).

Research Permission

The research project has been approved by both Duke University IACUC and a dedicated research committee at the DLC.

Research Training

All individuals collecting data are trained by the research staff from the DLC under the protocols approved by the IACUC.

Research Participation

All animals participated voluntarily in their home enclosure as an enrichment activity. Animals were free to withdraw from the testing area at any time. If the lemurs showed any signs of distress, the door to adjoining enclosures was opened immediately, reintroducing them to their group-mates.

Living Links, RZSS Edinburgh Zoo

Background

The 'Living Links to Human Evolution' Research Centre in RZSS Edinburgh Zoo has been designed as a scientific institution managed in collaboration between the Royal Zoological Society of Scotland, the University of St Andrews, and the Scottish Primate Research Group, which represents a consortium of primatologists at a number of Scottish Universities. The Centre has been created to facilitate behavioral, cognitive, and welfare-based research on naturalistically housed monkeys, at the same time introducing the zoo-visiting public to the science embodied in these enterprises, in an educational and even entertaining way.

Enclosures

Enclosures for the two mixed species groups mirror each other on either side of a central viewing platform and are named the 'West' and 'East' wings. Each wing includes an indoor squirrel monkey enclosure (5.5 m X 4.5 m X 6 m high), to which only the squirrel monkeys have access, an indoor capuchin enclosure to which both species have access (7 m X 4.5 m X 6 m high), and a large shared outdoor enclosure (approximately 900 m^2) to which both species have access. Between each pair of inner monkey enclosures is a research room, along each side of which is a set of two banks of cubicles, which form an entry and exit route for the monkeys, between their inner and outer enclosure. These cubicles can either be opened up to each other, or separated by transparent or opaque slides, thus providing a highly flexible research environment. Individual cubicles are 0.5 m^3, providing a run of 2 m long and 1 m high for each entire bank. The monkeys have permanent access to all areas of their enclosure except in inclement weather.

Animals

The Centre has housed two mixed species communities of common squirrel monkeys (*Saimiri sciureus*) and brown (tufted) capuchin monkeys (*Sapajus* sp.). These species co-habit in the wild. In the wild, capuchins have been shown to have a relatively small home range of 0.8 km^2 and squirrel monkeys of about 2 km^2. There were 35 capuchin monkeys (18 West, 17 East) and 30 squirrel monkeys (13 West, 17 East as of November, 2015). The two species live well together (Buchanan-Smith et al., 2013).

Diet

Subjects were never food or water deprived. The monkeys were fed on a rich diet of meat, eggs, vegetables, fruit, and monkey cereals. They were fed four times a day and also received regular food through enrichment devices and research rewards. See below for research procedures relating to food.

Ethical Approval

The research was approved by the School of Psychology & Neuroscience Ethics Committee of the University of St Andrews (the project entitled "Working memory in new world monkeys and great apes" was approved on 10/04/2018; no permit number was issued).

Research Permission

All projects must be approved by the research liaison officer (Living Links Team Leader employed by the zoo), the research director (employed by the University of St Andrews) and a Research Fellow based at the zoo (employed by an SPRG member university). Research projects must have ethical approval from

the lead researcher's institution and all researchers involved in the project need a Basic Disclosure Scotland before they can work at the zoo.

Research Training

All researchers undergo a relevant induction and training. This occurs after the project has received zoo approval and before the study begins. This may or may not be done before university ethical approval. However, studies cannot begin until the zoo receives evidence of university ethical approval. Training is led by a senior keeper. On their first day in the facility, researchers receive training relating to zoo Health and Safety, Zoo policies and theoretical training pertinent to their project. This is all based on an induction handbook which is emailed to researchers. Those working in the research rooms are then given additional practical training from a trained keeper. This involves about eight to 24 sessions (depending on training criterion being reached) where the keeper and the researcher work in the research rooms together. Researchers are trained by keepers to recognize individual monkeys, operate the sliders safely and identify behaviors in the monkeys. They are also given training on escapes and emergencies. Only once the keepers are satisfied that the researchers can work safely and can react appropriately to the animals' cues are they able to begin their study. Researchers also have an ID test with a keeper to ensure they know the identity of the monkeys.

Research Participation

Most monkeys have been habituated to remain in the research cubicles for research sessions in which they may be either by themselves or in various social configurations required for the particular research question under consideration. Participation is voluntary. A monkey is never forced to come into the research cubicles. Monkeys are isolated for up to 15 minutes, up to twice a day, four days a week. There is a clock in each research room and stop watches available to assist with time keeping. If the monkeys show any signs of distress, they are reintroduced to the group immediately. These signs of distress include ceasing participation, moving to the back of the cubicle and/or putting hands on the cubicle slides and/or emitting specific vocalizations.

There is no access to ad-libitum food and water in the cubicles, but monkeys are given regular food rewards during all research. The monkeys may be rewarded with sunflower seeds, nuts, raisins, dates, cereal, and mealworms. There are maximum allowances for these which have been decided by senior members of the husbandry team.

Other

The West squirrel monkey group is a breeding group, the other three populations (East squirrel monkey and East and West capuchin groups) are not breeding. Infants tend to stay on their mothers until they are about a year old. During this time the mother can still participate in research if they seem comfortable. Infants will not be isolated until there is an assessment by the keepers that the infants are regularly 'off' of their mothers and both the infant and the mother show no signs of distress during isolation.

Budongo Research Unit, RZSS Edinburgh Zoo

Background

The 'Budongo Research Unit' in RZSS Edinburgh Zoo is a research facility of the University of St Andrews in collaboration with the Royal Zoological Society of Scotland. Members of the Scottish Primate Research Group, a consortium of primatologists in several Scottish Universities, also participate in this collaboration. The Unit was created to promote the advancement of the scientific knowledge of the behavior, cognition, and welfare on naturalistically housed chimpanzees. It provides a unique opportunity to zoo visitors to observe researchers at work and learn about the latest developments in this area.

Enclosures

Enclosures include a large outdoor area (1985 m^2), three indoor 'pods' (total floor area 309 m^2, approx. 10 m traversable height in each), and an off-view 'beds' area (55 m^2), all interconnected by a tunnel system (30 m^2 approx.), and with access to climbing structures mounted on natural substrate (e.g., grass, dirt). Indoor pods provide varying levels of natural light and are air conditioned and temperature monitored. Moreover, enclosures possess visual barriers and an off-view area that allow individuals to retreat from other group members or the zoo visitors. Enclosures have access to free-flowing water and include a number of food enrichment devices provided to individuals daily. The group has access to all indoor areas during the night, which have multiple raised sleeping platforms at varying heights, and are provided with natural bedding materials (e.g., eucalyptus leaves, wood wool) and additional blankets/cardboard from which to make their nests. The main research area (30 m^2, 2.12 m high) is adjacent to the chimpanzee indoor enclosures. It consists of three adjoining research rooms that can be used as one large research area or split into three smaller research areas by hydraulic doors. Access to the research area and participation in research activities is completely voluntary. Moreover, there are multiple access routes from the indoor pods and tunnel system into the research area so that chimpanzees can enter and exit from multiple directions and never feel trapped in by other individuals of their group.

Animals

The Unit houses a group of 17 chimpanzees (*Pan troglodytes*), seven adult males, nine adult females and one juvenile male.

Diet

Subjects were never food or water deprived. The chimpanzees were fed on a varied diet of vegetables, fruit, nuts, seeds, eggs, and vegetation browse. They were fed four times a day and also received regular food through enrichment devices and research rewards. See below for research procedures relating to food.

Ethical Approval

The research was approved by the School of Psychology & Neuroscience Ethics Committee of the University of St Andrews (project entitled "Working memory in new world monkeys and great apes" was approved on 10/04/2018; no permit number was issued).

Research Permission

All projects must be approved by the research liaison officer (Living Links Team Leader employed by the zoo), the research director (employed by the University of York) and a Research Coordinator based

at the zoo (employed by the University of St Andrews). Research projects must have ethical approval from the lead researcher's institution and all researchers involved in the project need a Basic Disclosure Scotland before they can work at the zoo.

Research Training

On their first day in the facility, researchers receive training on health and safety as well as testing procedures. This information is compiled in the induction handbook that the research coordinator sends to each inductee. Once the project has been approved by the Zoo and the University ethics committee, the project may begin. Researchers are accompanied at all times by a keeper who provides practical advice and support during testing. Senior researchers may be approved to work on their own after a period of testing under keeper supervision.

Research Participation

Most chimpanzees entered the research area and participated in our research sessions. They entered as a group of varying composition or individually. Participation was strictly voluntary and they were free to leave the area at any time. If the chimpanzees showed signs of distress (e.g., whimpering) during the test, we terminated it immediately. Chimpanzees received food and/or fruit juice (diluted in water) during tests. Solid food included apples, raisins, cereal, and grapes. There were maximum food allowances for each of these items set by the senior members of the husbandry team.

Franklin and Marshall College

Background

The Franklin and Marshall College Vivarium in Lancaster, PA, USA, houses two separate family groups of capuchins (*Cebus/Sapajus apella)*. Research in the primate laboratory is voluntary on the part of the animals and non-invasive.

Animals

Eighteen capuchins in total were resident in the facility at the time of this research (n = 9 females, n = 9 males) and all participated in the study. No invasive research on primates has ever been allowed in the facility.

Ethical Approval/Research Permission

All research conducted on vertebrate animals at Franklin and Marshall College is reviewed and approved by the Institutional Animal Care and Use Committee (IACUC). Permission was granted for this study by the IACUC to the two involved PIs: Elizabeth Lonsdorf and Lauren Howard.

Enclosures

The two family groups are referred to as the 'F' and 'Y' colonies. The 'F' colony resides in a main housing enclosure, measuring 3.35 m wide X 8.50 m long X 3.05 m high, which can be divided into four smaller spaces via sliding mesh doors. Adjacent to the main enclosure is a wall of two rows of eight testing cubicles, each 0.91 m wide X 0.91 m long X 1 m high. The 'Y' colony resides in a main housing enclosure, measuring 3.16 m wide X 7.16 m long X 3.05 m high, which can be divided into three smaller spaces via sliding mesh doors. Adjacent to the main enclosure is a wall of two rows of six testing cubicles, each measuring 0.91 m wide X 0.91 m long X 1 m high. The two colonies are separated by an observation room of one-way mirrored safety glass, and therefore cannot see each other. See Lonsdorf et al. (2016) for a schematic. For this experiment, animals were tested individually in the testing cubicles after voluntary separation facilitated by positive reinforcement training.

Diet

Subjects were never food or water deprived. Meals of fresh produce and Mazuri Primate Diet (PMI Nutrition International St. Louis, MO) were scattered once daily and small quantities of cereal, fruit, nuts and mealworms were provided during routine husbandry training and for enrichment.

Research Training

Individuals who work with the monkeys undergo a formal course of training overseen by the Director of Animal Operations.

Research Participation

All methods were performed in accordance with the relevant guidelines and regulations of this committee and adhered to the American Society of Primatologists (ASP) Principles for the Ethical Treatment of Non Human Primates. No modifications were made to standard animal care routines. For testing sessions, individual subjects were brought into testing cubicles from the main housing enclosure

following standard positive reinforcement techniques. If any subject exhibited signs of stress, the session was terminated and repeated at a later time.

German Primate Center

Background

The German Primate Center is a research institute studying primates and currently houses 6 species of primates.

Animals

Lemurs. The ring-tailed and black-and-white ruffed lemurs were born in captivity and are housed in enriched outdoor and indoor cages at the German Primate Center. We tested 7 out of 8 ring-tailed lemurs belonging to two groups with 3 and 5 individuals. We tested 7 black-and-white ruffed lemurs that live in one group.

Macaques. The Cognitive Ethology Lab of the German Primate Center contributed data from 17 long-tailed macaques (*Macaca fascicularis*) to this study. The monkeys were born in captivity and lived in a social group of 36 individuals at the time of data collection (29 females, 7 males; age range: 1 - 30 years).

Enclosures

Ring-tailed lemurs are housed in indoor enclosures with one room of 3 X 3 m and two rooms of 3 X 3 m for each group and an outdoor enclosure of 29 X 20 m. The black-and-white ruffed lemurs are housed in two indoor enclosures of 4.4 X 3.6 m and an outdoor enclosure of 29 X 20 m. The two outdoor enclosures are next to each other, so that the two species see each other. All enclosures are enriched with tree trunks, ropes, and nets to climb, as well as wooden platforms to sit or lie on.

The long-tailed macaques have access to indoor and outdoor enclosures (49 m^2 and 141 m^2 respectively), which are equipped with various enrichment objects, wooden platforms, fire hoses, and a water basin during the warm months.

Diet

Subjects were never food or water deprived. Both species of lemurs were fed with a variety of vegetables, fruits and monkey chow. The macaques were fed their normal diet of monkey chow, fruits, and vegetables twice a day. Water was available ad libitum for all species.

Ethical Approval

Animal Welfare body of the German Primate Center.

Research Permission

Non-invasive studies have been reviewed and approved by the Niedersächsische Landesamt für Verbraucherschutz und Lebensmittelsicherheit and the Animal Welfare body of the German Primate Center.

The experiment was conducted in accordance with the German Animal Welfare Act and was approved by the ethics committee of the Animal Welfare Body of the German Primate Center (Permit Numbers: Lemurs E3-18_4-17, Long-tailed macaques: E3-18_9-17). Permission from the Lower Saxony State Office for Consumer Protection and Food Safety was not required (LAVES Document 33.19-42502-04).

Research Training

Lemurs. The experiments of this study were conducted by an experienced primate researcher working with lemurs for more than 20 years and a technical assistant.

Long-tailed macaques. The researchers were trained to work with the respective populations according to the local safety instructions.

Research Participation

Lemurs. All subjects participated voluntarily in the experiments and were never coerced. Subjects voluntarily entered the indoor testing cage, which was part of their home cage and could choose to end the testing by walking to the door, which then was immediately opened.

Long-tailed macaques. The monkeys were tested in a testing area adjacent to their indoor enclosure. The researchers offered participation in experiments by opening the doors between the indoor enclosure and the test area and monkeys were never forced to enter if they did not want to. During testing, the doors between the indoor enclosure and the test area were closed. The monkeys are used to being separated from their group for short periods of time when they participate in cognitive experiments. During testing, the subjects remained in visual and auditory contact with their group. The monkeys were familiar with the researchers and were used to interact with them for the purpose of training and testing.

Other

Lemurs: Adult females of both species are involved in a breeding program.

Gibbon Conservation Center

Background

The Gibbon Conservation Center hosts the largest population of gibbons in the United States of America.

Animals

The Gibbon Conservation Center currently hosts 5 different gibbon species including 12 Eastern Hoolock gibbons (*Hoolock leuconedys*, 5 females), 6 Pileated gibbons (*Hylobates pilatus*, 4 females), 8 Northern White Cheeked gibbons (*Nomascus leucogenys*, 3 females), 12 Javan gibbons (*Hylobates moloch*, 4 females) and 1 female Siamang (*Symphalangus syndactylus*). Only 3 Javan gibbons participated in the study. The gibbons from the Gibbon Conservation Center are not involved in any type of invasive research.

Enclosures

The gibbons located at the Gibbon Conservation Center live in large outdoor enclosures and they are usually grouped in pairs—sometimes together with their kin. The enclosures can be divided in two different compartments by sliding a metal mesh from outside. This mechanism allows us and the Gibbon Conservation Center staff to easily separate the gibbons prior to the study. Enclosures included multiple ropes and branches to facilitate brachiation from side to side of the enclosure.

Diet

Subjects were never food or water deprived. The gibbons were fed several times a day with a varied diet mostly composed of vegetables and fruit. The food provided as an incentive to participate in the study (blackberries) did not interfere with their feeding schedule.

Ethical Approval

The current research has been approved by the IACUC committee of the Gibbon Conservation Center (GCC) and complied with the rules of the IACUC office at University of California, San Diego.

Research Permission

Permission was given by an internal committee at the Gibbon Conservation Center.

Research Training

All researchers were previously trained according to the rules of the IAUCUC and the staff at the Gibbon Conservation Center.

Research Participation

Participation was voluntary for all individuals. Gibbons that lived with other group members were temporarily separated during the study period, when necessary.

Grastyán Translational Research Center

Background

The Grastyán Translational Research Center is a medium-sized NHP research facility associated with the University of Pécs. Currently we are working with 20 rhesus macaques in the facility.

As our research focuses on the cognitive behavioral domain, mainly on attention and short-term memory, our animals have various levels of experience in different touchscreen-based memory tasks. Specifically, they are well-trained in the delayed matching to sample (DMTS) and paired associates learning (PAL) paradigms using the commercially available MonkeyCANTAB test battery. The animals have been participating in minimally invasive behavioral pharmacology experiments using the above mentioned DMTS and PAL paradigms.

Animals

Seventeen male rhesus macaques participated in the experiment. Their age ranged from 5-14 years (mean ± standard deviation: 8.8 ± 2.8 years). The animals have not been involved in invasive research and never received major surgery or implants in their body. Most of the animals have been participating in behavioral pharmacology experiments and routinely received amnestic and/or cognitive enhancer agents using per os or systemic routes.

Enclosures

The animals are housed at the Grastyán Translational Research Center. In the vivarium they live in pairs in large home cages according to the legal requirements. Home cages are at least 200 X 100X 200 cm (length X width X depth) and fully comply with the 2010/63/EU Directive on animal experimentation. The home cages have two floors, where the second floor is made of wood. In the vivarium the illumination is close to the natural spectrum (there are also some windows), with a 12 hours light period followed by a 12 hours darkness period. The temperature is permanently kept at 24 +/- 2° C, with medium relative humidity (55 +/- 10%), and the air quality is renewed at a constant specified rate (10-20X/hr) by an air conditioning system.

Diet

Subjects were not food or water deprived for testing. The animals were fed with standard nutritionally complete dry pellets specifically designed for non-human-primates (Altromin Spezialfutter GmbH, Lage, Germany). This dry diet was supplemented daily with fresh fruits and vegetables. Animals were fed once per day, in the afternoons, following their daily testing sessions. As an incentive to participate in the study, a piece of peeled peanut, vegetable or raisin were used.

Ethical Approval

Two bodies have given ethical approval for the research:
   1) Local Animal Welfare Commitee, University of Pécs
   2) National Scientific Ethical Committee on Animal Experimentation (approval registration number is BA/35/62-5/2020).

Research Permission

The final approval by the National Scientific Ethical Committee stated that the research project entitled "Examination of Short-Term Memory in Non-Human Primates within the Frame of the

ManyPrimates International Project" could be conducted without detailed authorization from the national animal welfare authority. The justification states that the National Scientific Ethical Committee on Animal Experimentation of the Hungarian Government categorized the project as having 'no requirements for project license' (KA-2903, signed on January 28, 2020).

Research Training

Experimenters were trained to safely interact with the animals during everyday procedures and behavioral experiments for at least 8 months before conducting this study.

Research Participation

All animals showed interest in the task and were maximally willing to participate in the training and the tests. All but one animal finished the main test in one session. The animals did not show signs of noticeable distress or other signs of discomfort during the course of the study.

Other

The animals were not involved in any breeding program.

Heidelberg Zoo

Background

Zoo Heidelberg was founded in 1933 and currently houses 155 animal species, including gorillas and chimpanzees. The Zoo supports and conducts scientific studies on a regular basis.

Animals

The western lowland gorillas (*Gorilla gorilla gorilla*) housed at Zoo Heidelberg live in a social group of 4 animals. Three of the gorillas (1 male and 2 females) participated in the study, however, data collection could only be completed with the 2 females. The male stopped participating.

Enclosures

The enclosure of the gorillas consists of outdoor (160 m²) and indoor (393 m²) areas with connecting tunnels and several smaller compartments to separate the individuals if necessary. The enclosure is equipped with tree trunks, ropes, and nets to climb, stone and wooden platforms to sit or lie on. There are plastic barrels, wooden puzzle boxes and large plastic balls available for enrichment. The floor is covered with bark mulch, straw, and wood wool. Between the gorilla and chimpanzee enclosures there are windows allowing the apes to see each other and even interact on occasion.

Diet

The subjects were ever food or water deprived. They were fed with a variety of vegetables, leaf-eater pellets and browse and in addition received some cereals, nuts, puffed rice, etc. for enrichment. Water was available ad libitum.

Ethical Approval

Non-invasive studies are reviewed and approved by Heidelberg Zoo. This study was approved by the Heidelberg Zoo scientific department, consisting of Dr. Klaus Wünnemann, Director, Sandra Reichler, curator for mammals, conservation and research and Dr. Barbara Bach, zoo veterinarian. The scientific department is under continued supervision of the ethics committee of Heidelberg Zoo, headed by Dr. Klaus Zuber, Director of the Veterinary Department of the city of Heidelberg. We do not use permit numbers, therefore there is no number available for this study. Heidelberg Zoo is accredited by EAZA and the World Association of Zoos and Aquariums (WAZA).

Research Permission

Research conducted at Heidelberg Zoo complies with international and national standards and laws (e.g., Guidelines for the Treatment of Animals in Behavioural Research and Teaching published by the Association for the Study of Animal Behaviour) and institutional guidelines. Non-invasive studies are reviewed and approved by the curator and veterinarian of Heidelberg Zoo. Further IRB/IAUCUC approval was not necessary because no special permission for the use of animals in purely behavioral or observational studies is required in Germany (TierSchGes §7 and §8).

Research Training

An experienced primate researcher working with apes and monkeys for more than 10 years conducted the experiments. In addition, individuals working with the animals are trained by animal caretakers and receive further safety instructions.

Research Participation

All subjects participated voluntarily. No participation was ever coerced. Only subjects entering the testing area voluntarily participated in the experiments to ensure no stress is induced. No sliders were closed, and all animals could choose to end the testing and walk away whenever they liked.

Other

As the outdoor enclosure of the gorillas is going to be rebuilt soon, no breeding is taking place at the moment.

While the gorillas had experience using touchscreen computers, they had not participated in any manual task before. Therefore, before data collection could begin, we first had to train them to point to a food reward so that they could successfully indicate their selections during testing. To do so, we placed a small piece of fruit (e.g., grape or slices of pear) on the testing platform in view of, but out of reach of, the subject. When pointing at the food, either with their whole hand, a finger, or stick, they were handed the reward. Training sessions, each lasting no more than 10 minutes, were run every day with each subject until they were reliable.

Kristiansand Zoo

Background

Kristiansand Dyreparken (Zoo) is located 11 km east of Kristiansand (Norway). Kristiansand Zoo is the largest zoological institution in Norway and the only one in the country to house great apes. The institution houses in total eight species of primates.

Animals

Kristiansand Zoo houses a breeding group of 17 ring-tailed lemurs and three male black faced spider monkeys. All animals are captive born. Three female lemurs were tested as a group but only the two highest ranking females completed all trials. Only one spider monkey (the dominant male) completed all trials due to monopolization of the set up.

Enclosures

All tests were conducted in the sleeping rooms of the primates off-sight from the visitors. The lemurs had access to two indoor enclosures, one outdoor enclosure and a sleeping room approximately 5 x 3m. The lemurs were housed in two groups to prevent males from attacking the three newborns of the group. The two groups alternated indoor enclosures every day, as only one of the indoor enclosures allowed access to the outdoor enclosure. The spider monkeys had access to three connected sleeping rooms, each approximately 5 to 7 m long and 2 m wide. They were housed in a single group and had constant access to the indoor and outdoor enclosures and sleeping rooms (except during cleaning hours). Tests took place while the outdoor and indoor enclosures were being cleaned. Once tests and cleaning routines were finished, the subjects had access to both indoor and outdoor enclosures. Both indoor and outdoor enclosures were equipped with structural enrichment such as climbing frames, hose hammocks, artificial trees and logs, and bedding material. In the indoor and outdoor enclosures, the subjects had access to feeders where food was hidden every morning. The outdoor enclosure of the spider monkeys consisted of an island surrounded by a 5 m wide filled moat. The island included hanging bridges, huts and climbing frames. The lemurs' outdoor enclosure consisted of a fenced area of natural nordic forest.

Diet

Subjects were never food or water deprived. Diets at the zoo are designed by a veterinarian according to the species nutritional requirements. The lemurs' and spider monkeys' diets consisted of fresh fruits and vegetables, together with lower quantities of primate pellets and nuts. Food was provided twice a day, once in the morning and once in the afternoon. Feeding consisted in the scattering of food in the enclosures as well as the provision of food in localized areas. The rewards provided during testing were pieces of the subject's favorite fruit already present within the subject's diet.

Ethical Approval

The testing methodology for this project was approved by the Ethical Board for Scientific Research at Kristiansand Zoo, led by Rolf Arne, and by the Kristiansand Zoo Primate Project. The Ethical Board at Kristiansand Zoo does not issue permit numbers. Kristiansand Zoo is a member of EAZA and WAZA.

Research Permission

Research permission was granted by the Ethical Board for Scientific Research at Kristiansand Zoo, who internally decides which projects are conducted at the zoo. Decisions within this committee are made in collaboration with the veterinary staff and the animal keepers.

Research Training

The study was conducted by a doctoral candidate with experience in designing and performing cognitive tests with different primate species. The experimenter had conducted previous research at the testing institution and was familiar with the husbandry and safety procedures of the zoo.

Research Participation

Participation was completely voluntary and was conducted while the subjects were in their sleeping rooms as an enrichment activity. The only measure taken during the tests was to call the subjects' names if the subjects stopped participating to regain their attention.

Other

The lemurs included in this study are part of a breeding group and had infants with them (three) during the time of testing. Females with infants were kept separated from the rest of the group to prevent aggression from the males.

Kumamoto Sanctuary

Background

Kumamoto Sanctuary (KS) is the first and only sanctuary for chimpanzees and bonobos in Japan.

Animals

54 chimpanzees and 6 bonobos live in KS. 6 chimpanzees and 6 bonobos were included in this study.

Enclosures

Apes lived in an enriched environment with an outdoor compound (200-700 m^2) equipped with climbing structures and vegetation attached to indoor sleeping rooms (70-200 m^2). They lived in social groups consisting of 6-11 individuals.

Diet

Chimpanzees were given a variety of vegetables, fruits, nuts, and monkey chow three times a day, with additional enrichment items between the main meals. Water was available ad libitum. Neither food nor water was deprived for the purpose of experiments.

Ethical Approval

Wildlife Research Center, Kyoto University

Research Permission

Animal husbandry complied with international standards (the Weatherall report "The use of non-human primates in research") and institutional guidelines (Wildlife Research Center "Guide for the Animal Research Ethics"). The experimental protocols were approved by the Ethics Committee of the Wildlife Research Center, Kyoto University (WRC-2018-KS008A).

Research Training

Experimenters were trained to safely interact with apes and not to give subtle behavioral cues (e.g., gaze to the correct location) during the test for a minimum of three months.

Research Participation

All apes were tested in indoor sleeping rooms for each species. Upon testing, each individual ape was invited from the outdoor compound to the indoor sleeping room. Then, the door between the sleeping room and the outdoor compound was shut down to prevent other apes from coming in. All apes were willing to participate in the tests and did not show any strong stress behaviors (e.g., stress defecation) in this study. When chimpanzees stop participating in the experiments and/or show such strong stress behaviors (e.g., upon hearing conspecific fights outside), we let them out by opening the door.

Other

No breeding program is adopted in the sanctuary.

Insufficient info.

Lagos Zoo

Background

Lagos Zoo is situated in the south of Portugal, 12 km north-west of Lagos city. The zoo opened in 2000 and aims to provide an enriched enclosure specifically designed for each different species. The institution houses a total of fifteen primate species.

Animals

Lagos zoo houses two female and four male Emperor tamarins (*Saguinus imperator*) and three female golden handed tamarins (*Saguinus midas*) that participated in this experiment. Only one individual of each species completed the trials included in this study.

Enclosures

The individuals who participated in this study live in pairs (emperor tamarin) and in a group of two (golden-handed tamarin) in a fenced outdoor enclosure of approximately 56 m^2, with a sleeping boot of 2.25 m^2. Each cage is enriched with tree climbing trunks, vegetation that allows visual barriers, and wooden platforms.

Diet

Subjects were never food or water deprived. The two species of tamarins included in this study were fed twice daily, in the morning and afternoon. Their diet consisted of a variety of fresh fruits and vegetables, nuts, and dry fruits. The tamarins had fresh water available at all times. As an incentive for participating in the study, raisins, and their favorite fruits (such as grapes, strawberries, and oranges) were used.

Ethical Approval

The research complied with guidelines provided by the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (ETS n. 123). The research also adhered to the ASAB guidelines for the treatment of animals in behavioral research and teaching.

Research Permission

The permission to conduct this research was granted by the Lagos Zoo Board, considering the compliance with European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes and the ASAB.

Research Training

Data collection was conducted by a researcher after a week of collecting observational data on the two species of primates in order to be able to recognize the individuals accurately and for the individuals to be familiarized to the researcher's presence.

Research Participation

The individuals were never subject to any previous scientific experiment before and consequently the training was done gradually, first by presenting only a food reward, then one cup before a window, and

later the cup without the window. Each individual approached the set-up area and performed the test individually and participated voluntarily. During the training and experimental phase, each individual chose when to start, continue and leave. Training and experimental sessions did not last more than 10 to 15 minutes and were conducted once or twice a day.

Other

The individuals of this study were not part of any breeding or rewilding program.

Language Research Center, Georgia State University

Background

The Language Research Center (LRC) is an interdisciplinary research unit of the College of Arts and Sciences at Georgia State University. Although it was founded in 1981, its history begins a decade earlier in the ape-language research of founding Director Duane M. Rumbaugh and his collaborators. Historically, the LRC has housed and tested the cognition of bonobos, chimpanzees, orangutans, rhesus macaques, capuchin monkeys, and human children. At present, the LRC houses capuchin monkeys and rhesus monkeys.

Animals

The LRC contributed data from 21 capuchin monkeys.

Enclosures

The capuchin monkeys are housed in the CapLab, a separate facility with indoor (56 m^2) and outdoor (~135 m^2) areas and individual test cages for each animal, as well as group testing cages for observations of two or more animals. Additional (human-only) areas in this facility include areas for storage of cleaning supplies and personal protective equipment for staff, a kitchen area for storing and preparing food for the monkeys, and office space for record-keeping and experimental apparatus storage.

Diet

All animals are maintained on a veterinarian-approved daily diet that is supplemented with any food rewards used during cognitive testing. No food or water restriction is ever used with these monkeys.

Ethical Approval

This research was approved by the IACUC of Georgia State University. Georgia State University is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.

Research Permission

Research with animals at Georgia State University is approved by the IACUC committee under protocol A19042.

Research Training

All individuals are trained by the care staff and research staff at the Language Research Center under standard operating procedures approved by the IACUC.

Research Participation

All monkeys participate at their own choosing, voluntarily entering tests areas when offered the opportunity to engage in cognitive testing. No participation is ever coerced.

Other

Monkeys are monitored daily by research and care staff for psychological and physical wellbeing.

Lincoln Park Zoo

Background

Lincoln Park Zoo is located in Chicago, USA. The zoo is a leader in local and global conservation, animal care and welfare, learning, and science. A historic Chicago landmark founded in 1868, the not-for-profit Lincoln Park Zoo is a privately-managed, member-supported organization and is free and open 365 days a year. Currently, the zoo is home to around 200 animal species, including the swamp monkeys who participated in this study.

Animals

Two Allen's swamp monkeys (*Allenopithecus nigroviridis*) and three gorillas (*Gorilla gorilla gorilla*) housed at Lincoln Park Zoo were both tested as part of this study. The Allen's swamp monkeys were one male, Boko (aged 13 at time of testing) and one female, Kiden (aged 11 at time of testing). The swamp monkeys are housed together in the Helen Brach Primate House at Lincoln Park Zoo in a mixed-species exhibit with black-and-white colobus monkeys. The three gorillas were males: Azizi (age 14), Amare (age 12) and Mosi (age 11). They were housed together in an all-male group of four males in the Regenstein Center for African Apes.

Enclosures

The two swamp monkeys that participated in this study lived in a complex indoor exhibit and were provided with novel enrichment on a daily basis. The size of the exhibit was approximately 348.86 m^3 and includes multiple climbing structures and indirect natural light. In addition to their exhibit space are two off-exhibit "holding" areas (each appx. 17.49 m^3) where we ran all our test sessions. These inter-connected enclosures allowed us to voluntarily and briefly separate the monkeys from each other for training and test sessions. Each of these enclosures had a concrete floor, mesh sides, and elevated platforms.

The three male gorillas that participated in this study lived in a complex exhibit and were provided with novel enrichment on a daily basis. The indoor space featured a deep mulch floor, climbing structures, hammocks and visual barriers. The outdoor space had a grass floor, climbing structures, hammocks and visual barriers. They had access to their outdoor exhibit area whenever weather permitted. The total size of the indoor/outdoor exhibit was (1932 m^2). In addition to their large exhibit, where they spent most of their time (22-23 hours per day) there was also an off-exhibit "holding" area where they were moved to while animal care staff cleaned their exhibit. This is where testing took place.

Diet

Subjects were never food or water deprived. The swamp monkeys were fed with a variety of fresh fruit and vegetables daily, in addition to primate chow. The foods that were used as an incentive for participating in this study (a variety of fresh produce for the monkeys and peanuts for the apes) were reviewed and approved by veterinary and nutrition staff prior to the start of the experiment.

Ethical Approval

This study was approved by the Lincoln Park Zoo Research Committee (#2018-014), which is the governing body for all animal research at the institution. This research adhered to legal requirements in the United States of America and to the American Society of Primatologists' Principles for the Ethical Treatment of Nonhuman Primates.

The Lincoln Park Zoo Research Committee approved this study October 1, 2019. At the time the study was approved, the Lincoln Park Zoo Research Committee was chaired by the Zoo's Vice President

of Conservation and Science, Dr. Lisa Faust. The committee is comprised of representatives from the zoo's full-time senior research staff (typically 10+ PhDs, 2-4 MSs), senior animal care staff (one veterinarian, the VP of Animal Care and Horticulture, and the General Curator), and at least one representative from the Learning department; the Vice President of Communications is ex officio to this committee.

Research Permission

Participation in this study was approved by the Lincoln Park Zoo Research Committee.

Research Training

The researchers are employees of Lincoln Park Zoo and so have received extensive training working with primates and run daily cognitive testing sessions with primates housed at Lincoln Park Zoo. For this specific study, the researchers were trained not to provide verbal or visual cues to the monkeys during testing, nor to stare directly at the monkeys while they were working. There was no direct contact between experimenter and either monkey.

Research Participation

All participation by the monkeys was voluntary and, although not necessary for this study, a session would be stopped if a subject showed any sign of distress. Both monkeys participated in the study in their indoor holding area, which is off public exhibit. The monkeys were tested individually in their "holding" enclosures and keeper staff separated the monkeys for testing on behalf of the researchers. The gorillas were tested individually in their "holding" enclosures. The gorillas, who are part of an all-male group of four gorillas, are separated each morning by keeper staff as part of their typical husbandry routine. They are typically separated for no more than an hour and have continual visual, auditory and olfactory access with their group members while separated. All testing took place during this time.

Other

These swamp monkeys and gorillas are managed as part of their respective Species Survival Plan®.

While the swamp monkeys participate in regular positive reinforcement training sessions with keeper staff as part of their husbandry routine, they had not participated in any previous cognition studies before this one. Therefore, before data collection could begin, we first had to train them to point to a food reward so that they could successfully indicate their selections during test. To do that, we placed a food reward on the testing platform in view of, but out of reach of, a monkey. Using positive reinforcement training and shaping techniques, we trained the monkeys to reach for the food rewards. Training sessions, each lasting no more than 30 mins, were run every day with each subject until they were reliable. This lasted no more than three weeks.

While the gorillas had extensive touchscreen research experience, they had not participated in many manual tasks of cognition. Therefore, before data collection could begin, we first had to train them to point to a food reward so that they could successfully indicated their selections during test. To do that, we placed a peanut in a shell (the same reward we used during testing) on the testing platform in view of, but out of reach of, a gorilla. Using positive reinforcement training and shaping techniques, we trained the gorillas to reach for the food rewards. Some individuals were more successful, and more reliable, when given a short stick, and so we also provided them with a stick to point to the cups during testing. Training sessions, each lasting no more than 5 minutes, were run every day with each subject until they were reliable. This lasted no more than three weeks.

Łódź Zoo

Background

Zoological garden (Łódź Zoo) is open 365 days a year. During the experiment it was open 9 am - 6 pm.

Animals

6 lion-tailed macaques (*Macaca silenus*) live in the Łódź Zoo. 4 individuals (Naomi - age 14, Woolfie – age 29, Punio – age 15.5, Calineczka – age 8.5) were included in this study. All individuals (except Calineczka) participated in studies on the object permanence.

Enclosures

The monkeys lived in two social groups consisting of 3 individuals each. Each group was housed in an enriched indoor (brick building with some stumps and ropes) and outdoor enclosure (with climbing structures, stumps and ropes).

Diet

Subjects were never food or water deprived. The monkeys' food consisted of green vegetables, fruits (e.g., bananas, grapes, apples, strawberries).

Ethical Approval

The Local Ethics Committee for Animal Experimentation (permit number 3/ŁB11/2016)

Research Permission

Katarzyna Majecka (permit number 76/2015)
Dariusz Pietraszewski (permit number 60/2015)

Research Training

The Researchers have several years of experience conducting experiments with monkeys, horses and dogs, and have academic background for scientific work.

Research Participation

All individuals participate in the study voluntarily, and all animals can choose to end the testing and walk away whenever they like.

Other

Two individuals included in this study are part of a breeding program and had one infant with them during the time of testing.

Lund University Primate Research Station Furuvik

Background

Lund University Primate Research Station Furuvik (henceforth LUPRSF) is a collaboration between Lund University and Furuvik Zoo, which was formalized in 2007 with the stated purpose of studying the cognitive abilities of great and lesser apes. The station is located at Furuvik Zoo, near Gävle (Sweden) and is part of the Cognitive Science division (LUCS) at the Department of Philosophy. LUPRSF has purpose-built facilities for conducting non-invasive research with apes, being unique in this respect in Scandinavia, and gives access to seven chimpanzees, three Sumatran orangutans and two white-cheeked gibbons. No invasive research is allowed or possible at the station and all participation in experimental situations is voluntary. Further information and background can be found here: https://www.lucs.lu.se/primate-research-station-furuvik/

Animals

There are 7 chimpanzees (*Pan troglodytes*), 3 Sumatran orangutans (*Pongo abelii*) and 2 Northern white-cheeked gibbons (*Nomascus leucogenys*) at the station. This study involved the two adult orangutans: a male (Naong, 30 years) and a female (Dunja, 29 years). No invasive research is ever conducted at the station.

Enclosures

The orangutan enclosure at Furuvik Zoo includes two large, interconnected indoor spaces, as well as two outdoor islands which are interconnected by a bridge. Adjacent to the enclosure, there are three interconnected rooms used for research. Two of the rooms are provided with purpose designed mesh panels, where retractable tables can be mounted to allow e.g., choice tasks. The indoor areas have natural substrate as flooring, and are provided with logs, ropes, hammocks, etc. Such enrichment elements can also be found on the outdoor islands, although natural vegetation is dominating. The orangutans receive species-specific enrichment activities every day, based on a schedule that is rotated on a weekly basis.

Diet

Subjects are never food or water deprived. They received a varied species-appropriate diet that included vegetables, fruit, nuts, seeds, protein sources, as well as vitamin enriched pellets. To limit the intake of soluble sugars, fruit was given primarily as reinforcement during husbandry training. As an incentive for participating in the study, pieces of apple were used. Fresh water is provided *ad libitum* in the enclosure.

Ethical Approval

The research conducted at LUPRSF is non-invasive and falls under the definition of observational research with great and lesser apes housed at public zoos, as outlined in the Regulations and general recommendations for animal testing issued by the Swedish Board of Agriculture (SJV 2017:40). As such, only research that does not interfere or manipulate the animals' environment is possible at LUPRSF. No ethical permits are required to conduct such research at LUPRSF since the research is generally approved by Swedish law.

Research Permission

The suitability of research proposals is assessed by the scientific director of LUPRSF and the animal manager at Furuvik Zoo, in order to determine if research protocols infringe the conditions set by the law, i.e., if the protocols entail invasive procedures or manipulations of the animals' normal environment at the zoo.

Research Training

Experimenters received general training from the scientific director of the station and a keeper was present during testing. Subjects were familiar with the experimenter.

Research Participation

Subjects engaged voluntarily in testing by entering the experimental room, and were free to leave at any time. Dunja was accompanied by her infant. Both subjects have experience with choice tasks and indicate choices by pointing. A keeper was present during testing who monitored the procedure and distracted the animals not being tested at the time in order to reduce possible interference. For all testing carried out at LUPRSF, the duration and state of each animal is recorded. Specifically, we record how long an individual is engaged in a given task and whether the animal is in a calm, excited, stressed, or agitated state. These records are reported by the scientific director to the Swedish Board of Agriculture, which oversees matters related to animal welfare in Sweden, on a regular basis.

Other

The individuals involved in the study form a reproductive group. The female gave birth to an infant in November 2017.

Monkey Haven, Isle of Wight

Background

The Monkey Haven, Isle of Wight, houses the research facilities of the Macaque Cognition project from the University of Portsmouth (researchers involved: Marine Joly, Bridget Waller, and Jerome Micheletta).

Animals

Fifteen primate species, with a total of about 60 individuals, lived at the Monkey Haven during the study. Three rhesus macaques (*Macaca mulatta*) belonging to a group of 5 and 3 Barbary macaques (*Macaca sylvanus*), belonging to a group of 5 were included in this study.

Enclosures

The monkeys were housed in enriched enclosures, equipped with climbing structures and enrichment devices (food puzzles, boxes, etc.). They all have access to an outdoor compound. They lived in social groups consisting of 4-6 individuals. Furthermore, only those subjects voluntarily entering the area with the experimental setup participated in the study to ensure low stress levels.

Diet

Subjects were never food or water deprived. The monkeys were fed daily with assorted fruits and vegetables, nuts, seeds, and commercial monkey pellets.

Ethical Approval

Animal Welfare and Ethical Review Body (AWERB) of the University of Portsmouth

Research Permission

The research received approval by the AWERB (approval no. 4015B). All aspects of the study were covered by this ethical approval.

Research Training

All individuals were trained by the research staff from the Macaque Cognition Project, University of Portsmouth, under the protocols approved by the AWERB.

Research Participation

Cognitive testing required subjects to break from their social group and enter the testing area voluntarily. Only those subjects voluntarily entering the area with the experimental setup participated in the study to ensure low stress levels.

Other

Monkeys were monitored daily by research and care staff for psychological and physical wellbeing.

Paris Zoo

Background

The zoo, founded in 1934, is part of the National Natural History Museum (MNHN), a vast, multifaceted public institution committed to research, conservation, education and dissemination of knowledge and expertise. Thus, research is engrained in this zoo, and is carried out by its animal curators, students from affiliated universities, and interns from a variety of disciplines.

Animals

We had thought of testing the MP1 protocol with our two female Guyanan brown capuchins, but given their initial neophobic reaction, and plenty of other factors, such as availability of time, we've switched to woolly monkeys (*Lagothrix lagotricha*). They are 3 males of 13, 14.5 and 19 years old. None of the animals in our zoo have been involved in invasive research nor participated in cognitive studies.

Enclosures

Our woolly monkeys live in three indoor areas of 10, 10 and 20 m^2, and have free access to an outdoor enclosure of 330 m^2. Plenty of tree trunks, branches and ropes create a varying and complex three-dimensional space.

Diet

Subjects were never food or water deprived. Woolly monkey diet was approved and regularly revised by our vets, and it included fresh vegetables, fruit, and monkey chow. As an incentive for participating in the study, apples, carrots and raisins were used.

Ethical Approval

Cuvier committee (National Natural History Museum)

Research Permission

Our zoo is a member of EAZA and complies with all its rules and recommendations regarding animal housing and welfare. Research conducted at the National Natural History Museum needs to be approved by the Cuvier committee - which approved this protocol

Research Training

Only the primate curator has worked with the animals for this project. Additional participants have observed and might later be involved in the study.

Research Participation

Animals had no problems accepting the testing platform, and quickly learned to indicate their preference.
Animals normally have access to all three areas of the indoor space. Separating them for the duration of the experiments did not generate tensions, nor did they have to wait for their turn to participate. One session/individual was interrupted as he lost interest in the test, and wanted to play.

Other

This is a non-breeding group.

East China Normal University Primate Research Center

Background

The Primate Center of East China Normal University is located in Shanghai, China.

Animals

Over 40 rhesus macaques (*Macaca mulatta*) were housed at the center and 4 males were included in this study.

Enclosures

Indoor enclosure (2 m X 1.8 m X 1.8 m) for each rhesus macaque.

Diet

As an incentive for participating in the study. Rhesus macaques were fed with fruits, nuts, and monkey chow twice a day. Water was available *ad libitum*.

Ethical Approval

East China Normal University Institutional Animal Care and Use Committee (IACUC) gave ethical approval for this study.

Research Permission

East China Normal University Institutional Animal Care and Use Committee (IACUC) gave the permission to the MP1 study.

Research Training

Experimenters were trained to safely interact with monkeys and practices in accordance with regulations stipulated in the IRB.

Research Participation

They were tested in their home cages. Therefore, we consider their participation as voluntary. Animals could decide not to participate in the testing at any time, but they were generally very cooperative.

Other

The animals were previously housed in social housing (6 of them) but they were tested in this study in their individual home cages.

Primate Center of Strasbourg University

Background

The Primate Centre of Strasbourg University is a unique place in Europe, it stretches over seven hectares of wooded land, designed to house primates from nine different species. Geographical isolation and available space provide an ideal setting for studying primates in semi-free ranging conditions. Access to the site is restricted to professionals (i.e., researchers, students, veterinarians, animal caretakers) and state authorities (e.g., veterinary services, Research government department).

Animals

The Primate Center of Strasbourg University hosts an average of 700 animals under the supervision of a specialized team of 25 persons (ethologists, researchers, veterinarians, animal caretakers), backed by an animal welfare structure (SBEA) and an independent external ethics committee. Among the different social group of the Primate Centre, we worked on two groups of tonkean macaques (*Macaca tonkeana*) one of the groups was mixed in ages and sexes and composed by 26 individuals and the other one composed by five adult males. We also studied two groups (14 and 5 individuals) of rhesus macaques (*Macaca mulatta*), two groups (15 and 5 individuals respectively) of white-faced capuchins (*Cebus imitator*), one male pair of long-tailed macaques (*Macaca fascicularis*), one male pair of brown capuchins (*Sapajus apella*), two groups (4 and 2 individuals respectively) of green monkeys (*Chlorocebus sabaeus*), one group (5 individuals) of brown lemurs (*Eulemur fulvus*), one female pair of black lemurs (*Eulemur macaco*), and two Families (4 and 6 individuals) of common marmosets (*Callithrix jacchus*). However, animals participated voluntary to the experiment and thus not all groups' members could participate in the task. Thirty-six individuals were tested. None of the tested individuals is involved in invasive research.

Enclosures

Most of the tested groups live in semi-free ranging conditions in wooded parks (3000 to 5000 m^2) with permanent access to an indoor-outdoor shelter (10 to 20 m^2). The other smallest groups live in indoor-outdoor shelters (20 to 50 m^2).

Diet

Subjects were never food or water deprived. Individuals were fed with commercial primate pellets twice a day in the indoor shelter and received fresh fruit and vegetables once a week. Water was provided *ad libitum* in the indoor shelter. Each reward corresponded to 0.5% of the daily caloric intake, if the individuals were able to accumulate the maximum number of rewards in both trials it corresponded to 10% of the daily caloric intake.

Ethical Approval

In France, official approval is not requested for ethological studies based on animal volunteering. However, the Primate Centre (Silabe) has its own authorization for housing and breeding NHP (n°B6732636).

Research Permission

Research complied with the EU Directive 2010/63/EU for animal experiments

Research Training

Data were collected by PhD and master students with experience in designing and performing behavioral assessments with different primate species. Most of the experimenters had conducted previous research at different institutions and were familiar with the husbandry and safety procedures of the primate center. For those less experienced, permanent experienced researchers of the Primate Centre of Strasbourg University taught and trained them to safely interact with monkeys and not give them subtle behavioral cues.

Research Participation

All subjects participated voluntarily. They were free to stop and leave the experimental area at any time, even in the middle of a session or trial. Most of the individuals were tested in a special area for behavioral experiments (SAS). However, some individuals were tested in the park because they did not want to work in the SAS. All monkeys were highly motivated to participate.

Other

All tested subjects lived in multi-male/multi-female social groups or families (for marmosets) except two males brown capuchins living together without female and one of the Tonkean macaque living in a five males group.

Primate Station of the Department of Anthropology of the University of Zurich

Background

The Primate Station of the Department of Anthropology of the University of Zurich is situated at the Irchel campus in Zurich city. The station contains 31 indoor enclosures and 20 outdoor enclosures. In addition, the house contains heaters and humidifiers to simulate tropical conditions.

Animals

Seventy-two common marmosets (*Callithrix jacchus*) and 9 cotton top tamarins (*Saguinus oedipus*) were housed at Primate Station. A total of 16 common marmosets (3 family groups and 3 couples) were included in this experiment. No animal of Primate Station is involved in invasive research (all experiments are licensed at degree of severity = 0).

Enclosures

Their home enclosures are 2.5 m height X1.8 m width X 3.5 m depth and are enriched with several climbing structures such as natural branches and ropes. Each enclosure contains an infrared lamp and a sleeping box. The floor is covered with bark mulch. The enclosures of the two couples are slightly smaller (2 m height X 1 m width X 3m depth). Additionally, each group has access to an outdoor enclosure (2.5 m height X 1.8 m width X 3.5 m depth) which also contains structures as natural branches and ropes, climbing structures and soil with natural plants. During summertime and whenever the weather allows it (temperature above 10 degree), the animals have access to the outdoor part. These two parts are connected via a tunnel system.

Diet

Subjects were never food or water deprived. The marmosets were fed a porridge containing vitamin supplements in the morning and fresh fruits and vegetables around midday. Each afternoon, they received a protein snack such as mealworms or a piece of boiled egg. Regular diet was given at the same time and the same place as usual. Water is available ad libitum from water dispensers. The changes made in their routine living conditions were kept to a minimum and only if essential to the experiment.

Ethical Approval

Veterinary Office of the Canton of Zurich

Research Permission

Licensed by the Kantonales Veterinäramt ZH, license number ZH232/19; degree of severity = 0

Research Training

Experimenters receive theoretical and practical training held by Institute of Laboratory Animal Sciences (LTK) and it is accredited by FELASA (Federation of European Laboratory Animal Science Associations) and the Federation of Swiss Cantonal Veterinary Officers (VSKT). This course is approved by the Federal Food Safety and Veterinary Office as official education in Switzerland for experimenters and is also accredited by the FELASA according to the functions of the EU Directive (functions A, C, D and modules 10, 20, 21).

Research Participation

The animals participated on voluntary basis, i.e., they were not handled but trained to enter the experimental cages voluntarily. If they showed signs of distress (e.g., piloerection of the tail, trying to leave the test situation) the experiment stopped and was continued the next day.

Texas Biomedical Research Institute

Background

Texas Biomedical Research Institute in San Antonio, Texas, is home to the Southwest National Primate Research Center (SNPRC). Founded in 1941 by Thomas B. Slick, the institute aims to protect the global community through pioneering research and scientific discovery while maintaining the wellbeing of our animal models. In 1999, the SNPRC became the first new NPRC in more than 35 years. The 200-acre campus is home to over 2,200 primates. This includes the largest baboon population (around 1,000) in North America and the largest marmoset population (around 500) in the world dedicated to infectious disease and aging research.

Animals

Texas Biomedical Research Institute houses several primate species including macaques, vervets, capuchins, baboons, marmosets, and retired chimpanzees. For this study, the SNPRC contributed data from 6 individual baboons (*Papio anubis*). None of the animals tested in this study were involved in invasive research.

Enclosures

Study subjects for this project are housed in outdoor group housing of varying size. Animals have access to manipulable and structural enrichment at all times. The enclosures are equipped with climbing structures, perches, hanging barrels, visual panels, and cage toys. Nutritional enrichment is provided at least five times a week. Social groups consist of 2-12 individuals.

Diet

Subjects were never food or water deprived. Our animals received a veterinarian-approved diet. Supplement enrichment was provided no less than five times a week. Water was available *ad libitum*. The food rewards used as an incentive for participating in this study were counted as part of the animals' enrichment for that day.

Ethical approval

This research was approved by the IACUC of Texas Biomedical Research Institute. Texas Biomedical Research Institute is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.

Research Permission

Colony use protocols for both rhesus and baboons are approved by the IACUC committee. The procedures required for this study can be classified as occupational and nutritional enrichment according to our guidelines, which is covered under the animal's current protocols. A research request for this specific study was submitted and approved by the Texas Biomed IACUC.

Research Training

Research was conducted by an SNPRC behaviorist that worked closely with the tested subjects on a daily basis. Animals were trained by behavioral services staff at the SNPRC under standard operating procedures approved by the IACUC.

Research Participation

       Animal participation was completely voluntary. No participation was ever coerced. Individuals were not separated from conspecifics during study sessions, but willingly separated from the group at their own comfort level. If an animal showed signs of distress during the test, testing was halted immediately.

Other

       The animal's physical and psychological well-being was monitored daily by care staff.

Wolfgang Köhler Primate Research Center

Background

The Wolfgang Köhler Primate Research Center (WKPRC) is a project of the Max Planck Institute for Evolutionary Anthropology. It is operated in collaboration with the Leipzig Zoo. Research focuses on the behavior and cognition of the four species of great ape: chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*), orangutans (*Pongo pygmaeus*), and bonobos (*Pan paniscus*). Researchers and students from the University of Leipzig, and other universities around the world, conduct their research projects at the center guided by the personnel of the Center.

Animals

Eleven chimpanzees and 5 orangutans participated in this study. The chimpanzees sample consisted of 7 females (age range 16-42) and 4 males (age range 14-43); the orangutan sample consisted of 4 females (15-30) and one male (38). All individuals lived in social groups.

Enclosures

The indoor space featured climbing structures, wooden platforms, visual barriers. The outdoor space had a grass floor, climbing trees, wooden platforms, and visual barriers. They had access to their outdoor exhibit area whenever weather permitted. The total size of the indoor/outdoor enclosure was 430/4000 m^2 for the chimpanzees and 230/1680 m^2 for the orangutans. Both enclosures are equipped with shaking boxes and poking bins which allow the apes to engage in activities similar to their natural social and foraging behaviors (e.g., tool use).

Diet

Subjects are never deprived of food or water. The chimpanzees and orangutans were fed with a variety of fresh fruit and vegetables daily. The food rewards that were used as an incentive for participating in this study (grapes) were reviewed and approved by veterinary and nutrition staff prior to the start of the experiment. In addition, the apes regularly received special foods (e.g., chestnuts) that the keepers hid in certain areas of the enclosure to promote natural foraging activities. Other opportunities for special foraging activities (e.g., artificial termite mounds) were offered on a regular basis and special enrichment materials were provided for the apes every afternoon.

Ethical Approval

The study was ethically approved by the internal ethics committee of the Max Planck Institute for Evolutionary Anthropology and Leipzig Zoo. Members of the committee are: director of the WKPRC, Dr. J. Call, research coordinator at WKPRC, Dr. D. Hanus, zoo veterinarian, Dr. A. Bernhard, head animal keeper, F. Schellhardt, and assistant head animal keeper, M. Lohse. All researche at WKPRC is approved by this committee. No medical, toxicological, or neurobiological research of any kind is conducted at the WKPRC. Research was non-invasive and strictly adhered to the legal requirements of Germany. Animal husbandry and research comply with the "EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria", the "WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums" and the "Guidelines for the Treatment of Animals in Behavioral Research and Teaching" of the Association for the Study of Animal Behavior (ASAB).

Research Permission

Permission was given by an internal committee of the Max Planck Institute for Evolutionary Anthropology and the Leipzig Zoo (see above). No permit number was issued. Further IRB/IAUCUC approval was not necessary because no special permission for the use of animals in purely behavioral or observational studies is required in Germany (TierSchGes §7 and §8).

Research Training

The researchers are all employees of the Max Planck Institute. The keepers are all employees of Leipzig Zoo and so have received extensive training working with these animals.

Research Participation

All individuals were tested individually. Only those subjects voluntarily entering the area with the experimental setup participated in the study and a session would be stopped if a subject showed any sign of distress.

Other

In cooperation with the zoo, the Köhler Center supports efforts to conserve great apes, both in the wild and in captivity. The breeding program at the zoo is framed within the global strategy of the EEP, and some research focuses on the husbandry and care of great apes in captivity.

Wuhan Zoo

Background

Wuhan zoo is located in Wuhan, Hubei Province, China. It was first open to the public in 1985. It was founded and funded by the city government.

Animals

Francois's langurs (*Trachypithecus francoisi*) (2 females and 4 males).
No invasive research has ever been conducted at the zoo.

Enclosures

For Francois's langurs: indoor enclosure (32 m^2), outdoor enclosure (~35m^2). They live in a social group of 6, with free access between indoor and outdoor enclosures.

Diet

Subjects were never food or water deprived. The Francois's langurs daily diet included leaves, fruits, steamed buns and a small quantity of nuts. As an incentive for participating in the study, nuts and dates (both are part of their normal diet) were used.

Ethical Approval

The zoo's management approved their own caretakers to run study on their own animals.

Research Permission

See above.

Research Training

No systematic training. The animals were only trained (via positive reinforcement) to come when their names were called.

Research Participation

Testing happened in the indoor enclosure. The door to the outdoor enclosure was always open. Subjects could exit freely anytime.

Zoo Leipzig

Background

The Max Planck Institute for Evolutionary Anthropology conducts research on primate behavior and cognition in collaboration with the Leipzig Zoo. Besides the well-established Wolfgang Köhler Primate Research Center (WKPRC), which focuses on the four great ape species housed in "Pongoland", the Max Planck Institute also conducts research on other primates as part of the ManyPrimates project. This includes the following four species: Diana monkeys (*Cercopithecus diana*), Hamlyn's monkeys (*Cercopithecus hamlyni*), crowned lemurs (*Eulemur coronatus*), and golden lion tamarins (*Leontopithecus rosalia*).

Animals

Five Diana monkeys, four Hamlyn's monkeys, two golden lion tamarins, and three crowned lemurs participated in this study. The Diana monkey sample consisted of four females (age range 1-13) and one male (age 14); the Hamlyn's monkey sample consisted of four individuals housed in two separated groups: group 1) two males (3 and 10), group 2) one female (23) and one male (15); the golden lion tamarin sample consisted of one female (5) and one male (8); the crowned lemur sample consisted of two females (1 and 5) and one male (6). All individuals lived in social groups.

Enclosures

Three species, Diana monkeys, group 1 of the Hamlyn's monkeys, and crowned lemurs were housed in "Gondwanaland", a massive greenhouse with a tropical climate. Group 2 of the Hamlyn's monkeys were housed in rear animal husbandry, meaning that enclosures were not visible for visitors. During the time of the study, the golden lion tamarins were also housed in rear animal husbandry, as the female golden lion tamarin only recently moved to the Leipzig Zoo.

All three species in Gondwanaland had two enclosures: one serving as sleeping space/rear enclosure and one outside enclosure, from which the animals are visible to visitors. The sleeping spaces of all three species featured climbing structures and platforms. The times, when the animals were in the outdoor space typically correlated with the opening times of the Zoo.

The Diana monkeys shared their outdoor space with pygmy hippopotamuses (*Choeropsis liberiensis)*. It featured climbing structures, platforms, and visual barriers. The total size of sleeping space/outdoor space was 45 m^2 /230 m^2.

Group 1 of the Hamlyn's monkeys shared their outdoor space with Kirk's dik-diks (*Madoqua kirkii*). It featured climbing structures, platforms, and visual barriers. The total size of sleeping space/outdoor space was 27 m^2 /100 m^2.

The crowned lemurs shared their outdoor space with radiated tortoises (*Astrochelys radiate*). It featured climbing structures, platforms, and visual barriers. The total size of sleeping space/outdoor space was 25 m^2 /110 m^2.

Group 2 of the Hamlyn's monkeys had an indoor and outdoor space. It featured climbing structures and platforms. They had access to their outdoor area when weather permitted. The total size of indoor and outdoor space was 18 m^2 /25 m^2.

The golden lion tamarins were housed in an indoor and outdoor space not visible to visitors. Both indoor and outdoor space featured climbing structures and platforms. They had access to their outdoor area when weather permitted. The total size of sleeping indoor/outdoor space was 17 m^2 /10 m^2.

Diet

Subjects were never food or water deprived. The Diana monkeys and Hamlyn's monkeys were fed with a variety of fruit and other food, such as cooked meat, seeds, dried fruit, vegetables, and salad. The

Golden lion tamarins were mainly fed with a variety of fresh vegetables, small amounts of fruits, cooked meat, eggs, insects, and gum arabic as a surrogate for tree sap. In addition, the monkeys regularly received food that the keepers hid in certain areas of the enclosure to promote natural foraging activities and enrichment materials. The crowned lemurs were fed with a variety of fresh vegetables and other food, such as salad, insects, leaves, and small amounts of fresh fruit. All food rewards that were used as an incentive for participating in the study were reviewed and approved by veterinary and nutrition staff prior to the start of the experiments.

Ethical Approval

The study was ethically approved by the internal ethics committee of the Max Planck Institute for Evolutionary Anthropology and Leipzig Zoo. Members of the committee were: research coordinator at WKPRC, Dr. D. Hanus, zoo veterinarian, Dr. A. Bernhard, head animal keeper, F. Schellhardt, and assistant head animal keeper, M. Lohse. No medical, toxicological, or neurobiological research of any kind is conducted at Leipzig Zoo. Research was non-invasive and strictly adhered to the legal requirements of Germany. Animal husbandry and research comply with the "EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria", the "WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums" and the "Guidelines for the Treatment of Animals in Behavioral Research and Teaching" of the Association for the Study of Animal Behavior (ASAB).

Research Permission

Permission was given by an internal committee of the Max Planck Institute for Evolutionary Anthropology and the Leipzig Zoo (see above). No permit number was issued. Further IRB/IAUCUC approval was not necessary because no special permission for the use of animals in purely behavioral or observational studies is required in Germany (TierSchGes §7 and §8).

Research Training

The researchers are all employees of the Max Planck Institute. The keepers are all employees of Leipzig Zoo and so have received extensive training working with these animals.

Research Participation

Most subjects could be tested individually; only the young Diana monkeys had to be with their mother while testing. Sessions would be stopped if a subject showed any sign of distress.

Zoo Schönbrunn

Background

The complete name of the zoo is Tiergarten Schönbrunn and is located on the grounds of the Schönbrunn Palace in Vienna, Austria. It is the oldest continuously operating zoo in the world. The zoo houses a total of 10 primate species.

Animals

Tiergarten Schönbrunn houses three adult Bornean orangutans (*Pongo pygmaeus*; 2 females, one male).

Enclosures

The indoor area features climbing structures, wooden platforms, visual barriers. The outdoor space has a grass floor, climbing trees, wooden platforms, and visual barriers. Orangutans have access to their outdoor exhibit area whenever weather permits. The total size of the indoor/outdoor enclosure was 167/745 m². Both enclosures are equipped with enrichment devices such as shaking boxes and poking bins, which allow the apes to engage in activities similar to their natural social and foraging behaviors (e.g., tool use).

Diet

Subjects were never food or water deprived. The orangutans were fed on a varied diet of vegetables, fruit, yoghurt, seeds, and vegetation. They were fed three times a day and they received regular food through enrichment devices and research rewards.

Ethical Approval

Discussed and approved by the institutional ethics and animal welfare committee of the University of Veterinary Medicine Vienna (Austria) in accordance with GSP guidelines and national legislation (project entitled "The evolutionary origins of short-term memory in primates: a ManyPrimates project" was approved on 28/10/2019; permit number ETK-167/10/2019).

Research Permission

The project has been approved by the research and conservation manager and the deputy zoo director of the Tiergarten Schönbrunn. The research project has ethical approval from the lead researcher's institution.

Research Training

The research has been carried out by a trained animal caretakers of the Tiergarten Schönbrunn in close collaboration with the lead researcher who has experience in designing and performing cognitive tests with different primate species.

Research Participation

The participation in the test was completely voluntary. The test was conducted while the apes were in their sleeping rooms as an enrichment activity early in the morning before the apes got access to their indoor enclosure. All apes were willing to participate in the tests, and did not show any signs of stress.

Supplemental References

Buchanan-Smith, H. M. Griciute, J., Daoudi, S., Leonardi, R., & Whiten, A. (2013). Interspecific interactions and welfare implications in mixed species communities of capuchin (*Sapajus apella*) and squirrel monkeys (*Saimiri sciureus*) over 3 years. *Applied Animal Behaviour Science, 147*(3-4), 324-333.

Lonsdorf, E. V., Bonnie, K. E., Grim, M., Krupnick, A., Prestipino, M., Whyte, J. (2016). Social influences on the spread of arbitrary conventions among capuchin monkeys. *Behaviour, 153*, 633-654.

Macdonald, C. & Whiten, A. (2011). The 'Living inks to Human Evolution' Research Centre in Edinburgh Zoo: A new endeavour in collaboration. *International Zoo Yearbook*, *45*(1), 7–17.