First comparative approach to touchscreen-based visual object-location paired-associates learning in humans and a non-human primate.

Daniel Schmidtke[1,2*], Sandra Ammersdörfer[1,2], Marine Joly[1,3], Elke Zimmermann[1,2]

[1]Institute of Zoology, University of Veterinary Medicine Hannover, Hannover, Germany

[2]Center for Systems Neuroscience Hannover, Hannover, Germany

[3]Centre for Comparative and Evolutionary Psychology, Portsmouth, United Kingdom

[*]Correspondence: Daniel.Schmidtke@tiho-hannnover.de

Abstract

A recent study suggests that the touchscreen-based dPAL task on visual object-location paired-associates learning (PAL) allows effective translation from animal models to humans. Here, we adapted the task to a non-human primate (NHP), the grey mouse lemur, and provide first evidence for the successful comparative application of the task to humans and NHPs.

Young human adults reach the learning criterion after considerably less sessions (one order of magnitude) than young, adult NHPs, which is likely due to faster and voluntary rejection of ineffective learning strategies in humans and almost immediate rule generalization. At criterion, however, all human subjects solved the task by either applying a visuo-spatial rule or, more rarely, by memorizing all possible stimulus combinations and responding correctly based on global visual information. An error-profile analysis in humans and NHPs suggests that successful learning in NHPs is comparably based either on the formation of visuo-spatial associative links or on more reflexive, visually-guided stimulus-response learning. The classification in the NHPs is further supported by an analysis of the individual response latencies, which are considerably higher in NHPs classified as spatial learners.

Our results, therefore, support the high translational potential of the standardized, touchscreen-based dPAL task by providing first empirical and comparable evidence for two different cognitive processes underlying visual object-location paired-associates learning in primates.

Keywords

Paired-associates learning; Comparative cognition; Primates; Intelligence; Evolution

Introduction

One of the major challenges in evolutionary cognitive and biomedical research is the development of standardized cognitive testing procedures that allow for the comparative assessment of cognitive functions and malfunctions in specific domains. An increasingly popular approach to this problem is the adaptation of non-verbal, computerized tasks initially developed for human diagnostics to animals (e.g. (Bussey et al., 2012; Horner et al., 2013; Oomen et al., 2013)). While this strategy often led to valuable results in the past, the process of adaptation inherently involves the risks that task validity gets lost and that positive results in animal studies are potentially over-interpreted in an anthropomorphic sense. The former risk usually increases the longer the phylogenetic distance from humans - from great apes over smaller non-human primates (NHP) to rodents -, i.e. with an increasing need of protocol reductions due to both cognitive and physiological/motoric constraints of the chosen model species. Therefore, the opposite approach, to use standardized tasks from animal cognitive research to assess conserved cognitive functions in humans, was recently proposed (Nithianantharajah et al., 2015).

An animal protocol that likely fits this purpose is the dPAL task (the "d" in dPAL stands for "different") on visual object-location paired-associates learning. It was developed for rodent testing (Talpos, Winters, Dias, Saksida, & Bussey, 2009) and requires the subjects to procedurally learn to discriminate three different visual items (black-and-white shapes) and to associate each of them with one out of three possible locations on a touchscreen. At a given trial of the task, two of the three items are presented simultaneously, one as a rewarded item-place match and the second, *different* one as an unrewarded item-place mismatch. The dPAL task was found to be sensitive to pharmacological manipulations and targeted lesioning in rodents and involves hippocampus-based spatial processing and/or striatal stimulus-response learning (e.g. (Delotterie et al., 2015; C. H. Kim, Heath, Kent, Bussey, & Saksida, 2015; M. Kim, Kwak, Yu, & Kaang, 2016; Talpos et al., 2009)). In that respect, it differs from the closely related sPAL task (item-place match and item-place mismatch of a given trial are the "*same"* item presented as duplicate), in which sensitivity to pharmacological manipulations of the hippocampus seems to be missing (Talpos et al., 2009). A suggested reason for this insensitivity of the sPAL task to hippocampal manipulations is that it favours the utilization of a hippocampus-

independent conditional rule (Talpos et al., 2009). Further, a distinction has to be made between the dPAL task and the Human CANTAB PAL protocol: The latter requires the trial-unique formation and delayed retrieval of visuo-spatial paired associates and has very recently been validated against established neuropsychological measures of episodic memory (Lenehan, Summers, Saunders, Summers, & Vickers, 2016). This means that Human CANTAB PAL assesses a different construct in which both memory encoding and retrieval depend on medial-temporal structures (hippocampus proper and parahippocampal gyrus, respectively (de Rover et al., 2011); compare (Takahashi, Ohki, & Miyashita, 2002)). In dPAL, learning occurs procedurally, i.e. it is not a model for episodic or episodic-like memory in humans and animals, respectively. However, applying the dPAL task to mice and humans, it could be demonstrated that a human sample with disease-related Dlg2 deletions shows deficits in visuo-spatial paired associates learning parallel to those found in a sample of Dlg2 knockout mice (Nithianantharajah et al., 2015). Based on this finding, it was postulated that animal protocols, such as the dPAL task, could effectively be used to bridge the translational gap from animal models to humans, by assessing cognitive mechanisms that presumably are conserved across species (Nithianantharajah et al., 2015).

The first aim of our study was to train the grey mouse lemur (*Microcebus murinus*) in the highly standardized dPAL protocol to provide first comparative performance data from a NHP. Mouse lemurs are particularly suited for this purpose, as they are currently discussed as a natural, chronic NHP model of human brain-aging and Alzheimer's disease (AD) that could be used to complement the rodent models that are dominating the field (Joly, Ammersdorfer, Schmidtke, & Zimmermann, 2014; Schopf et al., 2014; Verdier et al., 2015; Verdier & Mestre-Francés, 2016): Mouse lemurs show age-related loss of functionality in motoric, sensory, and cognitive domains that is similar to the effects of senescence known from humans. In addition, some aged mouse lemurs *naturally* develop neuropathological features of an AD-like neurodegenerative disease, such as amyloid plaques, tau aggregation, and cerebral atrophy (for a concise overview see (Verdier & Mestre-Francés, 2016)). Thus, different from transgenic rodent models, mouse lemurs allow for research on disease development and, with maximum ages of up to 14 years in our colony, for longitudinal studies on long-term disease progression. Despite their potential as a natural model, a full mouse lemur genome

reference has recently been published (Mmur_3.0: GenBank assembly accession: GCA_000165445.3) and strategies for the establishment of a mouse lemur knockout library through a reverse-genetic approach are currently discussed (Ezran et al., 2017). Standardized, touchscreen-based tools for the assessment of appetitive conditioning learning and cognitive flexibility have recently been adapted to this species (Joly et al., 2014). A comparable protocol for the assessment of hippocampal integrity in mouse lemurs is currently missing, but urgently needed, as the hippocampal formation is among the brain areas that are the first to be affected by Alzheimer's disease (e.g. (Arnold, Hyman, Flory, Damasio, & Van Hoesen, 1991; Jack et al., 2000)). Apart from this biomedical aspect, mouse lemurs belong to a group of nocturnal primates that are often considered to represent an ancestral primate condition (Martin, 1990). Standardized, visuo-spatial PAL data from mouse lemurs would, thus, provide novel insights into the evolution of intelligence for both biomedicine and evolutionary anthropology. The second aim of the study was to additionally test a set of human subjects in dPAL for comparative reasons and to link results to those of verbal post-acquisition interviews to determine learning strategies in humans. Such data can help to identify comparable cognitive processes in humans and NHPs to further bridge the translational divide.

Materials and Methods

Research ethics. Animal testing was in accordance with the NRC Guide for the Care and Use of Laboratory Animals, the European Directive 2010/63/EU, and the German Animal Welfare Act. It was approved by the Animal Welfare Committee of the University of Veterinary Medicine and approved and licensed by the Animal Welfare Committee of the LAVES (ref. 33.12-42502-04-14/1454, 04/28/2014). All human subjects gave written informed consent to participating in the study and to the publication of their anonymized data. The used methods were in accordance with the current ethical guidelines of the German Psychological Society (DGP) and the American Psychological Association (APA) and approved by the Ethics Committee of the Hannover Medical School (ref. 2833-2015).

Subjects. We trained a total of twelve adult individuals of the grey mouse lemur (*M. murinus*; N_{female}=8; N_{male}=4; age range: 2-8 years) in the touchscreen-based dPAL protocol. Mouse lemurs were

born and kept at the breeding colony of the Institute of Zoology (University of Veterinary Medicine, Hannover; Landeshauptstadt Hannover: ref. 42500/1H, 01/15/2014; for details on animal housing see (Joly et al., 2014)). As intact vision plays a vital role in touchscreen-based cognitive testing, all NHPs considered for the study had been checked for ocular pathologies by a veterinarian ophthalmologist prior to testing (for methods compare (Dubicanac et al., 2016; Dubicanac, Radespiel, & Zimmermann, 2017; Dubicanac, Strueve, et al., 2017)). Only animals without any signs for impaired vision (e.g. prolonged pupillary reflex, corneal anomalies, uveitis, and advanced cataracts) were used as subjects. Furthermore, we tested twelve male, human adults (age range: 19-34 years) in the touchscreen-based dPAL protocol. Human subjects were recruited on the campus of the University of Veterinary Medicine. They were *naïve* as to the nature of the task.

Setup, stimuli, and general testing procedure. NHPs were tested on a daily basis with one session of 36 regular trials per animal and day. Testing took place during the first two hours of the animals' activity periods and in a room separate from the housing rooms using a customized version of the Bussey-Saksida Touchscreen Chamber (Model 80604, Campden Instruments LTD.; Fig. 1A) and a self-coded dPAL protocol running on ABET-II (Model 89505, Lafayette Instrument). The chamber had a symmetrically trapezoidal floor. The touchscreen was positioned at the long base (245 mm; front end) of the isosceles trapezoid, whereas a reward tray (**RT**, Fig. 1A), through which liquid rewards (apple juice) could be delivered, was positioned at the short base (130 mm; back end). The base-to-base distance was 330 mm and the volume accessible by the NHPs had a height of 100 mm. The touchscreen itself constituted the whole front wall of the chamber, but was covered by a black Perspex mask with three response windows (**1-3**, Fig. 1A), through which the NHPs had access to the screen and behind which the training items were presented (Fig. 1A). The response windows were square-shaped (45 x 45 mm) and separated from the adjacent window(s) by a distance of 20 mm. In general, only pictorial black-and-white items were used for training. For the actual dPAL, we chose the set of stimuli initially introduced by Talpos and colleagues (flower, airplane, and spider; Fig. 1B; (Talpos et al., 2009)) to allow highest possible comparability with preceding studies (e.g. (Bartko, Vendrell, Saksida, & Bussey, 2011; Nithianantharajah et al., 2015; Talpos et al., 2009)).

Humans were tested on a single day per subject and in several consecutive sessions with 36 regular trials per session. To keep comparability between species as high as possible, human subjects made their responses to a touchscreen from a disassembled Bussey-Saksida Touchscreen Chamber and were trained in a highly similar dPAL protocol (for minor differences see below). Both NHPs and humans were tested in the dark with the touchscreen being the only source of visible illumination. During the tests, the experimenter monitored the subjects' performance from an adjacent room.

dPAL in the NHP (M. murinus). Before the animals entered the dPAL task, they had to proceed through a 5-step autoshaping procedure in which they had learned to interact with the touchscreen chamber, i.e. to respond (by nose-poke or touch) to pictorial stimuli pseudo-randomly presented at one out of three possible positions on the touchscreen (**1-3**, Fig. 1A; for details of the autoshaping procedure compare (Joly et al., 2014)). In the dPAL task, animals had to learn to visually discriminate three pictorial stimuli (flower, airplane, and spider; Fig. 1B) and to associate each of them with a rewarded location on the touchscreen (see Video S1 for an example of a NHP performing the task). The dPAL stimuli were new to all subjects. The rewarded location for each stimulus was kept constant across trials and sessions (flower = "left"; airplane = "centre"; spider = "right"). At a given trial (for a flowchart overview see Fig. S1), two of the three stimuli were presented simultaneously, one at its rewarded location ($\mathbf{S^+}$), the other one at an "incorrect", unrewarded location ($\mathbf{S^-}$). The third response window was left blank ($\mathbf{S^-}$, Fig. 1B). A response to the S^+ led to a reward (15 µl apple juice). Reward collection triggered a 5 s inter-trial-interval (ITI), after which the next regular trial (new stimulus combination) could be initiated by revisiting the reward tray (**RT**, Fig. 1A). A response to one of the incorrect response windows (S^-) was signalled by a brief pure tone (2 kHz, 0.5 s) followed by a 5 s time-out and a 5 s ITI after which a correction trial (CT) could be initiated. During correction trials, the stimulus combination to which the animal previously had responded incorrectly was presented again and under the same conditions as a regular trial until the subject eventually responded to the S+. Within a complete session of 36 regular trials, the six possible stimulus combinations ($\mathbf{SC_1}$-$\mathbf{SC_6}$, Fig. 1B) were presented in a pseudo-randomized, balanced design. Animals were trained in the dPAL protocol until they reached a performance of 80% correct choices (correction trials excluded) in two

consecutive, complete sessions. A session ended after 36 completed regular (non-correction) trials or a maximum duration of one hour.

dPAL in humans. For human testing, the 5-step autoshaping was replaced by a short (10 trials) test session, in which the subjects were allowed to freely interact with the touchscreen. All subjects intuitively responded to the pictorial items presented pseudo-randomly at one of the three locations on the touchscreen and proceeded quickly through the test session. The task was slightly modified, as correct decisions were not physically rewarded, but signalled by a green checkmark presented at the center of the touchscreen (at a position above the response windows used for stimulus presentation). A red "x" was used to indicate incorrect responses to the subject. To initiate a new trial after the ITI had passed, subjects had to press a "next" symbol at the same position. All other protocol parameters (pictorial stimuli, sound of the reward pump, 2 kHz pure tone, ITI, time-out, number of trials/session, etc.) were exactly as in the NHP version. Between sessions, subjects had free access to beverages (water or caffeine-free lemonades) and sweets as compensation for their effort. After the learning criterion (80% correct choices in two consecutive, complete sessions) had been reached by a given participant, he was asked (I) for the rule that he believed was underlying the task and (II) whether he had changed his strategy during dPAL.

Statistics. All statistical analyses were conducted with R (R 3.2.3, 2015, The R Foundation for Statistical Computing). For descriptive statistics in Fig. 3B, mean and standard error of mean (±SEM) are presented, to allow direct comparison with published data from the rodent literature. To test individual error profiles for deviations from chance in the NHPs, we used χ^2-based Goodness of Fit statistics with Bonferroni correction for multiple testing. Only the last third of individual errors to criterion was analysed to minimize the noise in the data caused by initial trial and error learning and/or strategy switching. As the number of errors to criterion generally was high in the NHPs and even small deviations from chance became significant as a result of the sample size, we additionally used Cramer's V (φc; 'lsr' package in R) as an estimate of effect size. In humans, the number of errors to criterion generally was too small to use comparable inferential statistics. Median response and reward

latencies were compared between NHPs using asymptotic Wilcoxon signed rank statistics. The belonging effect sizes (r) were calculated from the Wilcoxon statistics as $r = z/\sqrt{N}$. Confidence intervals for individual medians are presented as 95% bootstrap confidence intervals based on 10000 bootstrap samples each. A possible correlation between the number of errors/correction trials and the number of self-reported assumed rules in human subjects was investigated using Spearman statistics.

Results

dPAL performance in the NHP (M. murinus). Based on their global performance (learning curves), the NHPs could be divided into three groups of individuals: (i) Animals belonging to the first group were excluded from the study after a minimum of 50 sessions, if they regularly failed to complete sessions of the dPAL task within the one-hour time limit (completion rates <25%; N=4; F_5-F_6, M_2-M_3; Fig. S2). This decision was made, since the learning criterion in dPAL requires the subjects to achieve a performance of at least 80% correct choices in two consecutive, *complete* sessions. In NHPs that regularly fail to finish the sessions within the time limit, this criterion cannot be applied, as it either is never reached or likely detects successful learning "too late". The inclusion of incomplete sessions was not an option: Performance measurements in these sessions often are biased towards low percentages, as subjects usually stop responding after incorrect trials. Also, in incomplete sessions with very low numbers of trials, extreme values of 0% or 100% regularly occur (compare Fig. S2), i.e. a criterion including incomplete sessions can easily be reached without actual learning. (ii) Animals of the second group eventually started to complete the dPAL sessions, but did not show any notable increase in task performance after a minimum of 120 sessions (≥4 months of daily training), i.e. performance fluctuated around chance level throughout the training (N=3; F_7-F_8, M_4; Fig. S3). (iii) Finally, animals of the third group eventually completed the dPAL sessions and reached the *a priori* learning criterion of 80% correct choices in two consecutive, complete sessions (N=5; one male: M_1; four females: F_1-F_4). F_1 reached this criterion after 2158 (+1454 correction trials = CT; 66 sessions; approx. 2 months) regular trials. F_2 and M_1 needed 2697 (+2011 CT; 76 sessions; approx. 2.5 months) and 2940 (+1642 CT; 85 sessions; approx. 3 months) regular trials, respectively (Fig. 2A, 3A). The two successful, aged adults (>7 years) reached the criterion after 5022 (F_4; +3398 CT; 150 sessions;

approx. 5 months) and 10207 (F_3; +6749 CT; 285 sessions; approx. 9.5 months) regular trials (Fig. 2B, 3A). Different from the dropouts, all successful NHPs showed a high tendency to complete the training sessions (completion rates ranging from 83.3 to 98.6%) and a continuous performance increase throughout the training (Fig. 2).

In order to learn more about the strategies used for task completion in the successful NHPs, we analysed the terminal errors (last third of the errors made; Tab. 1) of these five individuals. Error profiles were analysed separately for stimulus combination pairs with identical items (SC_1/SC_6, SC_2/SC_4, SC_3/SC_5; Fig. 4A) and stimulus combination pairs with identical S^+ (SC_1/SC_2, SC_3/SC_4, SC_5/SC_6; Fig. 4B). For the first case (stimulus combination pairs with identical items), the error distribution differed highly significantly from chance (33.3%; χ^2-test; Bonferroni corrected $p<0.01$) in M_1, F_3, and F_4 (Fig. 4A), but only in M_1 the belonging effect was of a medium size (Cramer's V $=\varphi c=0.229$) with an overrepresentation of terminal errors in SC_1/SC_6. All other effect sizes were small or neglectable ($\varphi c \leq 0.095$). For the second case (stimulus combination pairs with identical S+), the error distribution differed significantly from chance (33.3; χ^2-test; Bonferroni corrected $p<0.001$; $\varphi c \geq 0.19$) in all subjects, with an overrepresentation of terminal errors in SC_3/SC_4 and medium effect sizes ($\varphi c=0.19$ to 0.25) in F_1-F_4 (Fig. 4B). This difference in error profiles between the male NHP and the females was accompanied by differences in the individual, median response latencies (Tab. 2). M_1 showed a very low (1.78 s) median response latency (time interval between the onset of a given stimulus presentation and the touchscreen response by the animal) as compared to the other four individuals (F_1-F_4), for which the median response latencies were 1.5 to 2.5 times higher (2.59-4.37 s; compare Tab. 2 and Fig. S4A for a density histogram of the individual response latencies). The belonging median reward latencies, however, were low in all animals (M_1: 1.12 s; F_1-F_4: 0.92-1.32 s) and individual differences were much smaller than those observed for the response latencies (compare Tab. 2 and Fig. S4B for a density histogram of the individual reward latencies).

dPAL performance in humans. To investigate the range of possible strategies that can be used to reach the task criterion in dPAL, we tested a set of twelve human subjects that were later (during post-acquisition interviews) asked to verbally report the strategies they used. All humans reached the

criterion for task completion within 2-4 sessions, i.e. considerably faster than the other non-human mammals that have been tested in dPAL (i.e. rats (Talpos et al., 2009), mice (Bartko et al., 2011), and mouse lemurs; compare below), so far. Nevertheless, we could observe high inter-individual differences in the number of errors the human subjects made until criterion (correction trials in Fig. S5). These inter-individual differences were linked to differences in the number of possible rules the subjects rejected before they eventually found the correct one ($r_{Spearman}$=0.87, N=12, p=0.0002; compare Tab. S1). When asked for the suspected rule that underlies the task during individual post-acquisition interviews, 10 out of 12 subjects (S_1-S_7, S_{10}-S_{12}; Fig. 3A) correctly reported the object-location paired-associates rule underlying the paradigm and confirmed it as the one they consequently employed to reach criterion (Tab. S1). The two remaining subjects (S_8-S_9) reported to have memorized all possible stimulus combinations (SC_1-SC_6) and the belonging correct responses to solve the task, without recognizing a general rule (Tab. S1). Using this strategy, the latter two subjects belonged to the least effective human participants (Fig. 3A; compare Fig. S5 for the non-logarithmic graph). This allowed an analysis of their error profiles comparable to the NHPs, in which both subjects, just like NHP M_1, showed a clear bias for errors in trials with either SC_1 and/or SC_6 being presented (Fig. 4A).

Comparative data on dPAL learning dynamics in non-human mammals. For the sake of completeness, we compared the grouped learning curves of the successful, young NHPs (≤4 years) with the grouped learning curves reported for young rats (Talpos et al., 2009) and young mice (Bartko et al., 2011). The data reveals that learning performance in the NHP lies within the same range as learning performance in rodents (Fig. 3B). This comparison, however, is based on grouped learning dynamics alone and does not allow for a comparison of individual learning strategies involved in dPAL between the species. Comparative data on the error profiles in mice and rats, unfortunately, had not been available.

Discussion

The here-presented results are the first demonstration of a successful comparative application of the dPAL protocol in a non-human primate and humans. The study further provides a first analysis of possible solving strategies in humans and shows that humans can reach the task criterion using two

different strategies. They can solve the task either by a memorizing strategy, using the gross visual appearance of the presented stimulus combinations to learn the belonging correct responses, or by applying a spatial rule. As intended by the developers of the task (Talpos et al., 2009), the latter strategy includes the formation of visual object-location paired-associates (i.e. the mapping of different items onto absolute spatial positions) and was the one predominantly used in the human subjects. The finding of two distinct error profiles and response dynamics in the successful NHPs suggests a highly similar dissociation of two different solving strategies in mouse lemurs with a dominance of the spatial strategy, as we will discuss in the following paragraphs. We will start, however, with a discussion of the unsuccessful NHPs and suggestions on how their numbers can potentially be reduced in future studies on dPAL.

dPAL in the unsuccessful NHPs. Of the 12 tested NHPs, only 5 could successfully be trained to criterion. One possible interpretation of these results is that the behaviour shown by the successful animals is atypical for mouse lemurs. Based on the observations we made during the training and our experience with touchscreen-based testing in mouse lemurs from previous studies (e.g. (Joly et al., 2014)), however, we think that this is unlikely. Instead, we suggest that the observed ”failure” of some of the NHPs was due to protocol features that can readily be modified to potentially increase the number of successful learners without negative effects on construct validity: (i) For the unsuccessful NHPs that were excluded from the study after at least 50 sessions (N=4), as they regularly failed to complete sessions within the one-hour time limit, the main problem seemed to be a motivational one. We assume that the observed behaviour resulted most likely from the rule change between the last autoshaping sessions (every response to a pictorial stimulus is rewarded), which all subjects had regularly finished within the time-limit, and the actual dPAL task (only the item-place match is rewarded, whereas the item-place mismatch is not). This rule change inevitably entailed a sudden, considerable increase in the reward-work requirement that may have exceeded the motivational level of some of the subjects. As stated above, this does not mean that these subjects were unable to learn the task *per se*. It rather means that they would have needed a (much) higher number of absolute training days to improve dPAL performance and, more critically, that they could not reach the pre-

315 defined task criterion, which required them to complete the sessions. To possibly circumvent these
316 problems in the future, we suggest two alternative modifications to the protocol. In order to increase
317 the motivational level of the subjects at the time of the rule change to counteract the increase in
318 reward-work requirement, one could slightly reduce the subjects' food/caloric intake during the days
319 of the very first dPAL sessions. This modification would be easy to implement, but has ethical
320 implications that would have to be taken into consideration. It, therefore, could only be applied in a
321 very limited range. A second, less critical approach in terms of ethical considerations would be the
322 realization of a home-cage based training procedure with free access to the setup and a rolling criterion
323 instead of the session-based training. While being a more elaborate solution and probably more
324 difficult to implement, such a procedure would prevent that subjects have to be removed due to
325 unfinished sessions and likely reduce the absolute number of training days by increasing the amount of
326 daily training. (ii) Of the remaining three dropouts, which were removed after at least 120 sessions (all
327 successful young NHPs reached the training criterion between the 66^{th} and 85^{th} session), as their
328 performance still fluctuated around chance level (50%), two had a clear stimulus preference, which
329 they failed to overcome despite the correction procedure. The reason for failure in the third animal is
330 unclear. We think that the number of dropouts of this type can effectively be reduced by changing the
331 set of stimuli that constitute the different stimulus combinations from pictorial items to more
332 featureless items. While the "flower-plane-spider" set of stimuli was the one routinely used in dPAL at
333 the time the here-reported experiments were conducted (e.g. (Bartko et al., 2011; M. Kim et al., 2016;
334 Talpos et al., 2009); Fig. 1B) and was chosen to guarantee a maximum degree of comparability, we
335 would consider using the set of line stimuli introduced by Kim and colleagues (C. H. Kim et al., 2015)
336 to minimize the negative effect of stimulus preferences on learning in future studies.
337
338 *dPAL in the successful NHPs.* Within the successful NHPs, we could distinguish two different error
339 profiles and response dynamics: One NHP (M_1) showed an error-profile with an overrepresentation of
340 errors in SC_1/SC_6 trials among the last third of individual errors made (Fig. 4A). This pattern suggests
341 a stimulus-response strategy in M_1, as such a strategy would, just like in the humans who memorized
342 all possible stimulus combinations (compare below), either be based on differentiating the gross visual

appearance of the stimulus combinations or on recognizing the sequence (e.g. from left to right) of individual items. Both the gross visual appearance and the sequence of individual items are highly similar in SC_1/SC_6 (Fig. 4A), so that this stimulus combination pair can be expected to be the most difficult to learn for individuals using a stimulus-response strategy. The remaining NHPs (F_1-F_4), on the other hand, showed a bias towards the pair of stimulus combinations in which the S^+ was presented in the centre position (SC_3/SC_4; Fig. 4B). This pattern is indicative for a spatial strategy in F_1-F_4, as SC_3/SC_4 is the most challenging stimulus combination pair in terms of spatial processing: Firstly, the rewarded S^+ is in the centre position. The respectively corresponding item-place mismatches (S-), therefore, change position from left (in SC_3) to right (in SC_4, Fig. 4B). In all other stimulus combination pairs with identical S^+ (SC_1/SC_2, SC_5/SC_6), the corresponding item-place mismatches are always on the same side (Fig. 4B). Secondly, in both SC_3 and SC_4 the corresponding item-place mismatches are directly adjacent to the S^+ (Fig. 4B), i.e. this stimulus combination pair has an increased difficulty in terms of location discrimination as compared to SC_1/SC_2 and SC_5/SC_6 with a larger spatial distance between S+ and S- in one stimulus combination per pair (Fig. 4B). In line with this, those of our NHPs that were classified as spatial learners (F_1-F_4) showed an increased (factor: 1.5 to 2.3) mean error frequency in stimulus combinations with narrow spatial distance between S^+ and item-place mismatch (SC_1, SC_3, SC_4, SC_6) as compared to the mean error frequency in stimulus combinations with large spatial distance between S^+ and item-place mismatch (SC_2, SC_5). This was not the case in the NHP that was classified as a non-spatial learner (M_1; factor: 0.9; compare Tab. 1). In further support of the classification of M_1 as a non-spatial learner and F_1-F_4 as spatial learners, M_1 showed a very low median response latency as compared to F_1-F_4 (Tab. 2 and Fig. S4A), while the median reward latency of M_1 was well within the range of the other subjects (Tab. 2 and Fig. S4B). This means that the special position of M_1 in terms of response latencies was not due to a motoric or motivational advantage of M_1, but that the short response latencies in M_1 are likely to mirror fast, reflexive decisions for a given response window based on visual stimulus appearance alone, whereas the significantly longer response latencies in the remaining individuals are likely to be caused by longer lasting decision-making processes that take both stimulus identity and position into account.

It is intriguing, that the NHP classified as non-spatial learner was the male individual among those subjects who reached criterion, whereas the spatial learners all were females. It is well described in the literature on both humans and rodents that internal levels of gonadal steroids can modulate learning strategies. Female rats that were tested in a continuously rewarded spontaneous alteration task in a Y-maze and a food finding task in a T-maze, for example, showed a bias towards spatial strategies at pro-oestrous (high levels of ovarian steroids), whereas female rats at oestrous preferentially used response strategies in the same tasks (Korol, Malin, Borden, Busby, & Couper-Leo, 2004). Comparably, in humans, women tested in a virtual navigation task at high progesterone levels during the mid/late luteal phase also showed a bias towards spatial strategies (Hussain, Hanafi, Konishi, Brake, & Bohbot, 2016). While it is unclear, whether the distribution of spatial and non-spatial learners between the sexes we observed is pure coincidence, we can likely exclude the possibility that a specific oestrous state has led to a bias towards a spatial strategy within our female subjects: Grey mouse lemurs have seasonal reproductive patterns and, in captivity, start cycling approximately one month after a change from an artificial short-day period (LD 10:14; at our colony from October to January) to a long-day period (LD 14:10; February to September). During the long-day period, female mouse lemurs are polyoestrous with 3-4 cycles per year that can vary between 42 and 68 days in length (Radespiel & Zimmermann, 2001; Wrogemann, Radespiel, & Zimmermann, 2001). During the subsequent short-day period, grey mouse lemurs are anoestrous. Of the four female NHPs that reached the task criterion, three started the dPAL training during the long-day period (F_2: 23rd of February; F_3: 7th of March; F_4: 30th of March). Due to the long training durations, each of these female subjects went through at least one full oestrous cycle before reaching criterion. The fourth female NHP (F_1) started and finished the dPAL training during the short-day period (18th of October – 22nd of December) while being anoestrous.

A second effect on dPAL in mouse lemurs that is indicated by our data is an age effect. While the sample size of successful NHPs is too low for inferential statistics, the clear difference in the number of trials needed to reach the criterion between young and aged adults (for age classification compare (Joly et al., 2014)) suggests that the number of trials needed to reach the criterion of the task increases with increasing age. Since all NHP subjects had been checked for impaired vision by an

ophthalmologist prior to testing and only individuals with good vision were included in the study, the performance difference between young and aged adult NHPs cannot be explained by visual deficits of the aged subjects. If an age-effect on dPAL in mouse lemurs could be verified in a future study, this would highly support their value as a *natural* and *chronic* NHP model of human brain-aging and Alzheimer's disease, as which they are currently discussed (Joly et al., 2014; Verdier et al., 2015; Verdier & Mestre-Francés, 2016), especially because a standardized task that assesses hippocampal malfunctioning is currently lacking in mouse lemurs.

Comparative data on dPAL in non-human mammals and humans. The comparison of our results with published data from the rodent literature on dPAL showed that, in terms of learning dynamics, mice, rats, and mouse lemurs are comparably slow and that humans are considerably faster in reaching the task criterion. While one would normally also expect the tested NHP to outperform the rodents, the fact that rodent performance is actually *en par* with that of the tested NHP corroborates the postulation that successful completion of the dPAL task in mammals relies on conserved cognitive mechanisms (Nithianantharajah et al., 2015) (e.g. hippocampus-based spatial learning and/or striatum-based stimulus-response learning (Delotterie et al., 2015; C. H. Kim et al., 2015; M. Kim et al., 2016; Talpos et al., 2009); compare below). The humans, on the other hand, had several decisive advantages over the animals tested in the task: Firstly, while animals must learn to discriminate the three items that constitute the different stimulus-combination pairs, this step probably is obsolete in the human subjects, due to the pictorial nature of the items (flower, airplane, spider). This is an additional reason why we would recommend the utilisation of more abstract, featureless stimuli (e.g. (C. H. Kim et al., 2015)) for future studies. Secondly, the human subjects had the advantage of a fast, voluntary rejection of ineffective strategies as well as almost immediate rule generalization once they had learned the first object-location paired-associate by trial and error. These abilities, however, require the conscious expectation of the existence of an underlying rule, which is probably unique to humans.

Nevertheless, the post-acquisition interviews revealed that humans can also use two different strategies to solve the dPAL task, a spatial one, in which each item is mapped to an absolute, correct location, and a memorizing strategy, in which the correct response is learned for each stimulus-

combination pair without the necessity for absolute spatial mapping. The two human subjects who self-reportedly chose the latter strategy could, just like the NHP M_1, be identified based on their error profiles: Towards the end (last third of individual errors made), these non-spatial learners also showed a clear bias for errors in trials with either SC_1 and/or SC_6 being presented (Fig. 4A). Both subjects reported that they were confused by the visual similarity between SC_1 and SC_6, as it consists of identical stimuli ("flower" and "spider") presented in the same spatial order ("flower" on the left side, directly adjacent to the "spider" on the right side), but differs in the belonging correct locations. For rodents, a comparable analysis of the error profiles had not been available in the literature. However, a dissociation between two possible learning strategies in dPAL has also been proposed for mice and rats, based on pharmacologic and exitotoxic lesioning studies conducted in these species ((Delotterie et al., 2015; C. H. Kim et al., 2015; M. Kim et al., 2016; Talpos et al., 2009); compare below).

The translational value of dPAL. As stated in the introduction, the Human CANTAB PAL and the animal dPAL model different, though possibly related, psychological constructs: The human protocol requires the tested subjects to recall the position of several visual stimuli on a computer display on a trial unique basis and after a brief delay between stimulus presentation and retrieval (Sahakian et al., 1988). In the here-described dPAL protocol, the task is acquired incrementally and in each trial a choice has to be made between a simultaneously presented object-location match *vs.* an object-location mismatch (e.g. (Horner et al., 2013)). Due to the lack of both trial uniqueness and the delayed response, the dPAL paradigm cannot be seen as a model for episodic or episodic-like memory in humans and animals, respectively. Nevertheless, clinical evidence for the translational value of dPAL was provided by Nithianantharajah and colleagues who showed parallel cognitive deficits in mice and humans (human CANTAB PAL) with genetic perturbations of the Dlg2 gene (Nithianantharajah et al., 2013). There are two possible explanations for this finding: (I) Even though Human CANTAB PAL and dPAL model different psychological constructs, performance in both depends on a common cognitive component that is equally affected in humans and mice with Dlg2 mutations. If this is true, the most obvious common link between the two paradigms would be the necessity to retrieve combined visual and spatial information, a cognitive function that has also been shown to be

hippocampus-dependent in the absence of trial uniqueness and delay in rats using a non-CANTAB protocol (Yoon, Seo, Kim, & Lee, 2012). (II) Human CANTAB PAL and dPAL do not rely on homologue cognitive functions, but there is overlap in the brain areas involved in performing both tasks (e.g. the hippocampal formation). Which one of the two options is true is difficult to test. To avoid this general dilemma, i.e. translational problems resulting from species specific adaptations of protocols initially designed for humans, a recently suggested approach is the utilization of identical, highly controlled, touchscreen-based cognitive tasks designed for animal testing across all species, including humans (Nithianantharajah et al., 2015). Indeed, it was shown that the same parallel cognitive deficits as in the preceding study (Nithianantharajah et al., 2013) also became apparent when both mice *and* humans with Dlg2 gene mutations were tested in dPAL (Nithianantharajah et al., 2015). The authors argue that using the identical task across species, from mice to humans, highly increases construct validity as it is more likely that under these conditions the involved cognitive processes are adequately homologous between different mammalian species, though probably more basal and conserved as those assessed by more complex protocols designed for humans. Our study supports this suggestion and the suitability of the dPAL protocol for broadly comparative research, as it shows for the first time that the highly standardized dPAL protocol can directly be used to train a nocturnal NHP (*M. murinus*) in object-location paired-associates learning. Learning performance in mouse lemurs was not different from that reported in rodents (Bartko et al., 2011; Talpos et al., 2009), suggesting that dPAL is based on conserved cognitive mechanisms that need to be further specified: From the rodent literature, it is known that post-acquisition dPAL performance in rats is impaired after the pharmacologic manipulation of the dorsal hippocampus using glutamatergic antagonists (Talpos et al., 2009) or parenteral, systemic administration of NMDA antagonist or indirect dopamine agonist (Talpos, Aerts, Fellini, & Steckler, 2014). In mice, genetic manipulation of the glutamatergic system (TNiK-/-) revealed impaired dPAL acquisition in knockouts as compared to wild type mice (Coba et al., 2012) and lesions to the dorsal hippocampus led to impaired dPAL performance both during and after acquisition (C. H. Kim et al., 2015). A second study using excitotoxic lesioning of the hippocampus only found post-acquisition impairments in dPAL, whereas acquisition was severely disrupted in animals with striatal lesions (Delotterie et al., 2015). The most likely explanation for the

fact that post-acquisition lesioning of the dorsal hippocampus robustly affects dPAL performance in rodents (Delotterie et al., 2015; C. H. Kim et al., 2015; M. Kim et al., 2016; Talpos et al., 2009) while acquisition is not (Delotterie et al., 2015; Talpos et al., 2009) or only mildly (C. H. Kim et al., 2015) affected by hippocampus lesions is that intact animals acquire the task in a hippocampus-dependent manner (hence the profound effect of post-acquisition lesioning) but switch to alternative (equally effective) learning strategies (e.g. stimulus-response learning) if lesioning has occurred prior to acquisition (Delotterie et al., 2015; C. H. Kim et al., 2015). Our results are in line with the idea that two alternative strategies can be used for successful dPAL acquisition, as the error profiles in mouse lemurs either show biases towards stimulus combination pairs with increased object similarity (SC_1/SC_6) and short response latencies (N=1; indicative for a stimulus-response strategy) or for stimulus combination pairs with increased demands on spatial processing (S_3/S_4) and long response latencies (N=4; indicative for a spatial strategy). They further show that the spatial strategy, i.e. the mapping of objects onto locations, is the one predominantly used for successful task completion in both mouse lemurs (N=4; 80%) and humans (N=10; 83%).

Conclusion

Our study showed that the dPAL task on visuo-spatial paired associates learning originally designed for rodent testing (Talpos et al., 2009) can be used successfully to train a non-human primate as well as humans. This lays the foundations for the assessment of standardized paired-associates learning across different primate species to track cognitive changes over aging in order to match physiological profiles and behaviour in a comparative approach. To reach criterion, both the tested NHPs and humans seem to rely on one of two alternative cognitive strategies: Most of the subjects tested here used a strategy that includes spatial processing (suggesting a high construct validity), as intended by the developers of the task (Talpos et al., 2009). Much fewer subjects used a strategy including visually-guided stimulus response learning. This is in accordance with neurobiological models of dPAL in rodents, in which an involvement of hippocampal and striatal regions in dPAL was found in pharmacologic and exitotoxic lesioning studies (Delotterie et al., 2015; C. H. Kim et al., 2015; M. Kim et al., 2016; Talpos et al., 2009). Therefore, our findings support the recent postulation that dPAL in

mammals relies on conserved cognitive mechanisms (Nithianantharajah et al., 2015). By demonstrating for the first time that the protocol can be applied to a promising NHP model of human brain-ageing, they further suggest that the highly standardized dPAL (and similar animal-testing protocols) may function as unique tool for biomedical research and its translation to the clinic, due to its broad applicability from rodents over NHPs to humans. Such a "reverse" approach to cognitive testing can contribute to explore mechanisms of disease progression and novel therapeutic avenues in psychiatric diseases, but will also provide novel insights into the evolution of intelligence in mammals in general.

Competing interests: We have no competing interests.

Author contributions: DS, SA, MJ, and EZ conceived and designed the study. DS, SA, and MJ conducted the experiments in the NHPs, DS conducted the experiments in humans. DS coded the NHP version of the dPAL protocol and designed and coded the human dPAL protocol. DS performed the data analysis and wrote the first draft of the manuscript. SA, MJ, and EZ participated in writing. DS prepared the figures. EZ contributed financial support, reagents, materials, and analysis tools.

Funding: This study is part of a project which has received funding from the European Community's 7th Framework Programme (FP7/2007–2013) under grant agreement n° 278486 acronym "DEVELAGE" (http://www.develage.eu/index.html).

Acknowledgements: We thank Jennifer Wittkowski, Marko Dubičanac, Annette Klaus, Philipp Hohenbrink, and May Hokan for helping to collect parts of the data. Further, we thank the animal keepers: Lisa Früh, Iris Grages, and Johanna Samtlebe. We would further like to thank the anonymous reviewers for their helpful and constructive comments that contributed to improving the final version of the manuscript.

References:

Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R., & Van Hoesen, G. W. (1991). The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb Cortex, 1*(1), 103-116.

Bartko, S. J., Vendrell, I., Saksida, L. M., & Bussey, T. J. (2011). A computer-automated touchscreen paired-associates learning (PAL) task for mice: impairments following administration of scopolamine or dicyclomine and improvements following donepezil. *Psychopharmacology, 214*(2), 537-548. doi:10.1007/s00213-010-2050-1

Bussey, T. J., Holmes, A., Lyon, L., Mar, A. C., McAllister, K. A. L., Nithianantharajah, J., . . . Saksida, L. M. (2012). New translational assays for preclinical modelling of cognition in schizophrenia: the touchscreen testing method for mice and rats. *Neuropharmacology, 62*(3), 1191-1203. doi:10.1016/j.neuropharm.2011.04.011

Coba, M. P., Komiyama, N. H., Nithianantharajah, J., Kopanitsa, M. V., Indersmitten, T., Skene, N. G., . . . Grant, S. G. N. (2012). TNiK is required for postsynaptic and nuclear signaling pathways and cognitive function. *Journal of Neuroscience, 32*(40), 13987-13999. doi:10.1523/Jneurosci.2433-12.2012

de Rover, M., Pironti, V. A., McCabe, J. A., Acosta-Cabronero, J., Arana, F. S., Morein-Zamir, S., . . . Sahakian, B. J. (2011). Hippocampal dysfunction in patients with mild cognitive impairment: a functional neuroimaging study of a visuospatial paired associates learning task. *Neuropsychologia, 49*(7), 2060-2070. doi:10.1016/j.neuropsychologia.2011.03.037

Delotterie, D. F., Mathis, C., Cassel, J. C., Rosenbrock, H., Dorner-Ciossek, C., & Marti, A. (2015). Touchscreen tasks in mice to demonstrate differences between hippocampal and striatal functions. *Neurobiology of Learning and Memory, 120*, 16-27. doi:10.1016/j.nlm.2015.02.007

Dubicanac, M., Joly, M., Struve, J., Nolte, I., Mestre-Frances, N., Verdier, J. M., & Zimmermann, E. (2016). Intraocular pressure in the smallest primate aging model: the gray mouse lemur. *Vet Ophthalmol*. doi:10.1111/vop.12434

Dubicanac, M., Radespiel, U., & Zimmermann, E. (2017). A review on ocular findings in mouse lemurs: potential links to age and genetic background. *Primate Biology, 4*, 215-228. doi:10.5194/pb-4-215-2017

Dubicanac, M., Strueve, J., Mestre-Frances, N., Verdier, J. M., Zimmermann, E., & Joly, M. (2017). Photoperiodic regime influences onset of lens opacities in a non-human primate. *PeerJ, 5*, e3258. doi:10.7717/peerj.3258

Ezran, C., Karanewsky, C. J., Pendleton, J. L., Sholtz, A., Krasnow, M. R., Willick, J., . . . Krasnow, M. A. (2017). The Mouse Lemur, a Genetic Model Organism for Primate Biology, Behavior, and Health. *Genetics, 206*(2), 651-664. doi:10.1534/genetics.116.199448

Horner, A. E., Heath, C. J., Hvoslef-Eide, M., Kent, B. A., Kim, C. H., Nilsson, S. R. O., . . . Bussey, T. J. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nature Protocols, 8*(10), 1961-1984. doi:10.1038/nprot.2013.122

Hussain, D., Hanafi, S., Konishi, K., Brake, W. G., & Bohbot, V. D. (2016). Modulation of spatial and response strategies by phase of the menstrual cycle in women tested in a virtual navigation task. *Psychoneuroendocrinology, 70*, 108-117. doi:10.1016/j.psyneuen.2016.05.008

Jack, C. R., Jr., Petersen, R. C., Xu, Y., O'Brien, P. C., Smith, G. E., Ivnik, R. J., . . . Kokmen, E. (2000). Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology, 55*(4), 484-489.

Joly, M., Ammersdorfer, S., Schmidtke, D., & Zimmermann, E. (2014). Touchscreen-Based Cognitive Tasks Reveal Age-Related Impairment in a Primate Aging Model, the Grey Mouse Lemur (*Microcebus murinus*). *PLoS one, 9*(10), e109393. doi:10.1371/journal.pone.0109393

Kim, C. H., Heath, C. J., Kent, B. A., Bussey, T. J., & Saksida, L. M. (2015). The role of the dorsal hippocampus in two versions of the touchscreen automated paired associates learning (PAL) task for mice. *Psychopharmacology, 232*(21-22), 3899-3910. doi:10.1007/s00213-015-3949-3

Kim, M., Kwak, C., Yu, N.-K., & Kaang, B.-K. (2016). Optimization of the touchscreen paired-associate learning (PAL) task for mice and its dorsal hippocampal dependency. *Animal Cells and Systems, 20*(5), 229-236. doi:10.1080/19768354.2016.1221855

Korol, D. L., Malin, E. L., Borden, K. A., Busby, R. A., & Couper-Leo, J. (2004). Shifts in preferred learning strategy across the estrous cycle in female rats. *Horm Behav, 45*(5), 330-338. doi:10.1016/j.yhbeh.2004.01.005

Lenehan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., & Vickers, J. C. (2016). Does the Cambridge Automated Neuropsychological Test Battery (CANTAB) distinguish between cognitive domains in healthy older adults? *Assessment, 23*(2), 163-172. doi:10.1177/1073191115581474

Martin, R. D. (1990). *Primate Origins and Evolution: A Phylogenetic Reconstruction*. London: Chapman and Hall.

Nithianantharajah, J., Komiyama, N. H., McKechanie, A., Johnstone, M., Blackwood, D. H., St Clair, D., . . . Grant, S. G. N. (2013). Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nature Neuroscience, 16*(1), 16-U37. doi:10.1038/nn.3276

Nithianantharajah, J., McKechanie, A. G., Stewart, T. J., Johnstone, M., Blackwood, D. H., St Clair, D., . . . Saksida, L. M. (2015). Bridging the translational divide: identical cognitive touchscreen testing in mice and humans carrying mutations in a disease-relevant homologous gene. *Scientific Reports, 5*, 14613. doi:10.1038/Srep14613

Oomen, C. A., Hvoslef-Eide, M., Heath, C. J., Mar, A. C., Horner, A. E., Bussey, T. J., & Saksida, L. M. (2013). The touchscreen operant platform for testing working memory and pattern separation in rats and mice. *Nat Protoc, 8*(10), 2006-2021. doi:10.1038/nprot.2013.124

Radespiel, U., & Zimmermann, E. (2001). Dynamics of Estrous Synchrony in Captive Gray Mouse Lemurs (Microcebus murinus). *International Journal of Primatology, 22*(1), 71-90. doi:10.1023/A:1026466015443

Sahakian, B. J., Morris, R. G., Evenden, J. L., Heald, A., Levy, R., Philpot, M., & Robbins, T. W. (1988). A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain : a journal of neurology, 111*, 695-718. doi:10.1016/j.neuropharm.2011.04.011

Schopf, C., Zimmermann, E., Tünsmeyer, J., Kästner, S. B. R., Hubka, P., & Kral, A. (2014). Hearing and Age-Related Changes in the Gray Mouse Lemur. *JARO: Journal of the Association for Research in Otolaryngology, 15*(6), 993-1005. doi:10.1007/s10162-014-0478-4

Takahashi, E., Ohki, K., & Miyashita, Y. (2002). The role of the parahippocampal gyrus in source memory for external and internal events. *Neuroreport, 13*(15), 1951-1956. doi:10.1097/00001756-200210280-00024

Talpos, J. C., Aerts, N., Fellini, L., & Steckler, T. (2014). A touch-screen based paired-associates learning (PAL) task for the rat may provide a translatable pharmacological model of human cognitive impairment. *Pharmacology Biochemistry and Behavior, 122*, 97-106. doi:10.1016/j.pbb.2014.03.014

Talpos, J. C., Winters, B. D., Dias, R., Saksida, L. M., & Bussey, T. J. (2009). A novel touchscreen-automated paired-associate learning (PAL) task sensitive to pharmacological manipulation of the hippocampus: a translational rodent model of cognitive impairments in neurodegenerative disease. *Psychopharmacology, 205*(1), 157-168. doi:10.1007/s00213-009-1526-3

Verdier, J. M., Acquatella, I., Lautier, C., Devau, G., Trouche, S., Lasbleiz, C., & Mestre-Frances, N. (2015). Lessons from the analysis of nonhuman primates for understanding human aging and neurodegenerative diseases. *Frontiers in Neuroscience, 9*, 64. doi:10.3389/Fnins.2015.00064

Verdier, J. M., & Mestre-Francés, N. (2016). *Microcebus murinus* – a unique primate for modeling human brain disorders, including Alzheimer's disease and bovine spongiform encephalopathy. In S. M. Lehman, U. Radespiel, & E. Zimmermann (Eds.), *The Dwarf and Mouse Lemurs of Madagascar* (Vol. 73, pp. 161-173). Cambridge: Cambridge University Press.

638 Wrogemann, D., Radespiel, U., & Zimmermann, E. (2001). Comparison of Reproductive
639 Characteristics and Changes in Body Weight Between Captive Populations of Rufous and Gray
640 Mouse Lemurs. *International Journal of Primatology, 22*(1), 91-108.
641 doi:10.1023/A:1026418132281
642 Yoon, J., Seo, Y., Kim, J., & Lee, I. (2012). Hippocampus is required for paired associate memory with
643 neither delay nor trial uniqueness. *Learning & Memory, 19*(1), 1-8.
644 doi:10.1101/lm.024554.111

645

646

Tables:

Table 1: Number of terminal (last third) errors separated into individual stimulus combinations and NHPs. The ratio given in the rightmost column represents the mean number of errors in stimulus combinations with narrow spatial distance between S^+ and item-place mismatch (SC_1, SC_3, SC_4, and SC_6) divided by the mean number of errors in stimulus combinations with large spatial distance between S+ and item-place mismatch (SC_2 and SC_5).

NHP	SC_1	SC_2	SC_3	SC_4	SC_5	SC_6	Ratio
M_1	118	132	42	47	70	139	0.856
F_1	74	52	107	128	56	68	1.745
F_2	149	45	127	198	75	77	2.296
F_3	231	190	469	611	372	378	1.503
F_4	126	108	253	222	117	307	2.018

Table 2: Individual median reward and response latencies of the NHPs. Confidence intervals (CI) are presented as percentile bootstrap confidence intervals based on 10000 bootstrap samples per median. Effect sizes were calculated from Wilcoxon statistics as $r = z/\sqrt{N}$ with M_1 being the reference and F_1-F_4 being compared to M_1. The Response/Reward ratio was calculated by dividing the response latency of a given animal by the reward latency of the same animal.

NHP	Median response latency [s] (95% CI)	Effect size: r	Median reward latency [s] (95% CI)	Effect size: r	Response/Reward ratio
M_1	1.777 (1.776, 1.827)	-	1.117 (1.116, 1.117)	-	1.59
F_1	2.945 (2.893, 3.023)	0.80	0.916 (0.915, 0.964)	-0.33	3.22
F_2	2.589 (2.539, 2.640)	0.74	0.965 (0.965, 0.966)	-0.34	2.68
F_3	3.554 (3.503, 3.603)	0.78	1.066 (1.066, 1.067)	-0.66	3.33
F_4	4.370 (4.267, 4.469)	0.83	1.321 (1.320, 1.321)	-0.40	3.31

Figures:

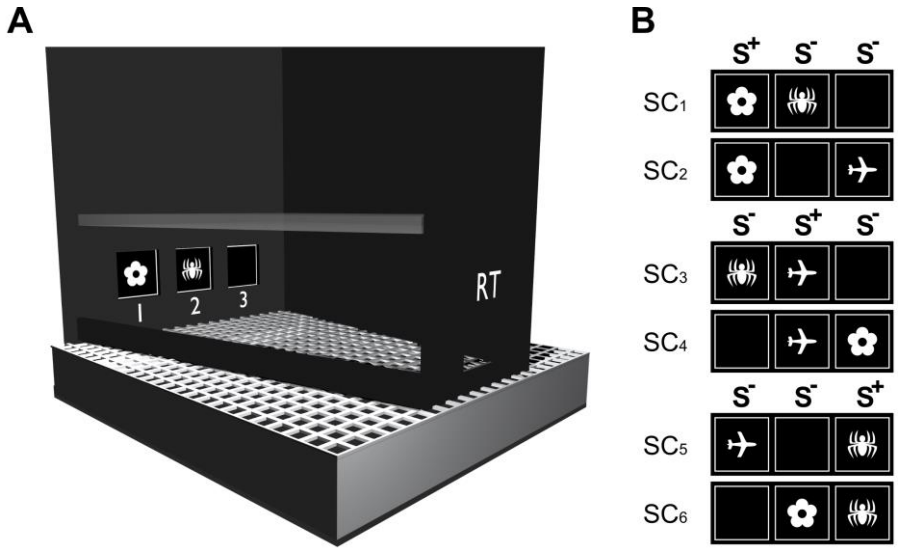

Fig. 1: Experimental setup and procedure. A Schematic drawing of the automated Bussey-Saksida Touchscreen Chamber (left sidewall and reward pump removed); **1-3** response windows 1-3; **RT** entrance to the reward tray; to keep the animals from climbing, the chamber height was limited to 10 cm using a translucent Plexiglas lid. **B** Stimulus combinations (**SC_1-SC_6**) that were used for dPAL training.

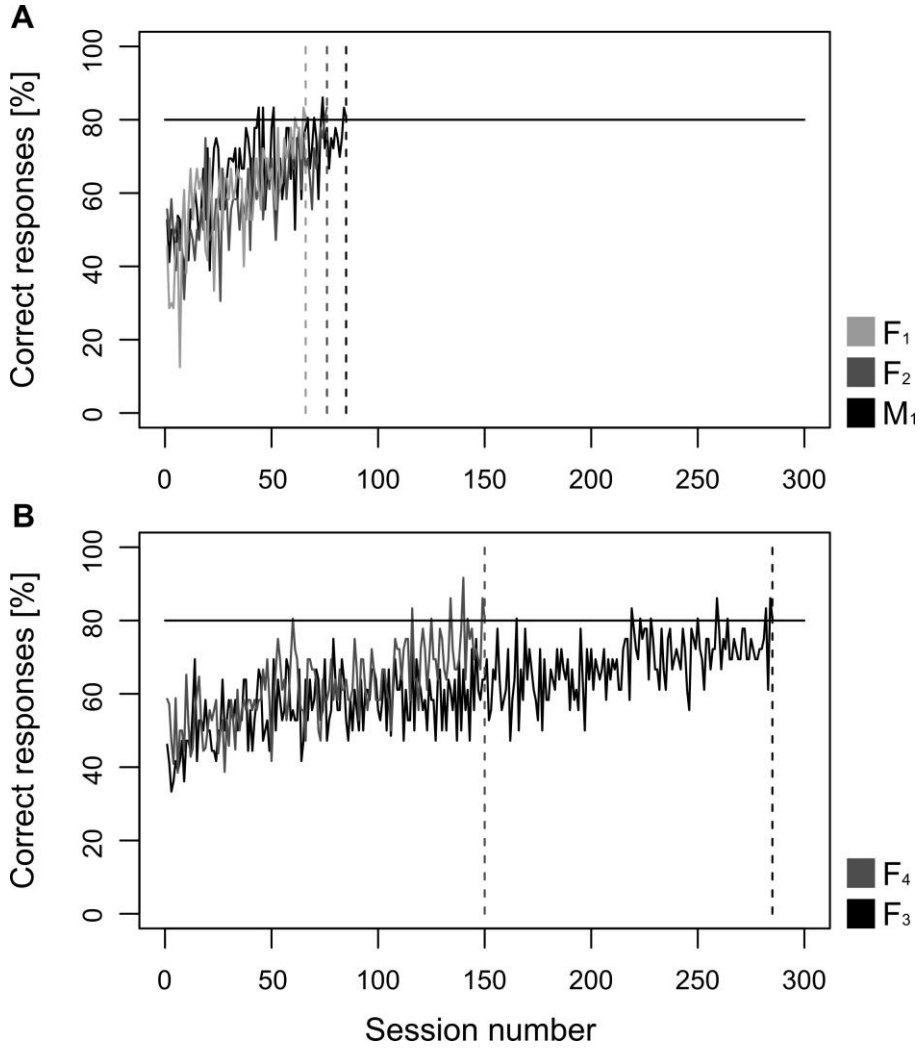

Fig. 2: Individual learning curves of the successful mouse lemurs. A Individual learning curves of the three young (<4 years) adults (F_1, F_2, M_1). **B** Individual learning curves of the two aged (>7 years) adults (F_3, F_4). **A**-**B** The black, solid, horizontal line indicates the 80% learning criterion that had to be reached in two consecutive, complete sessions in order to finish the task; the vertical, dashed lines indicate the sessions at the end of which the criterion was reached by the respective individuals.

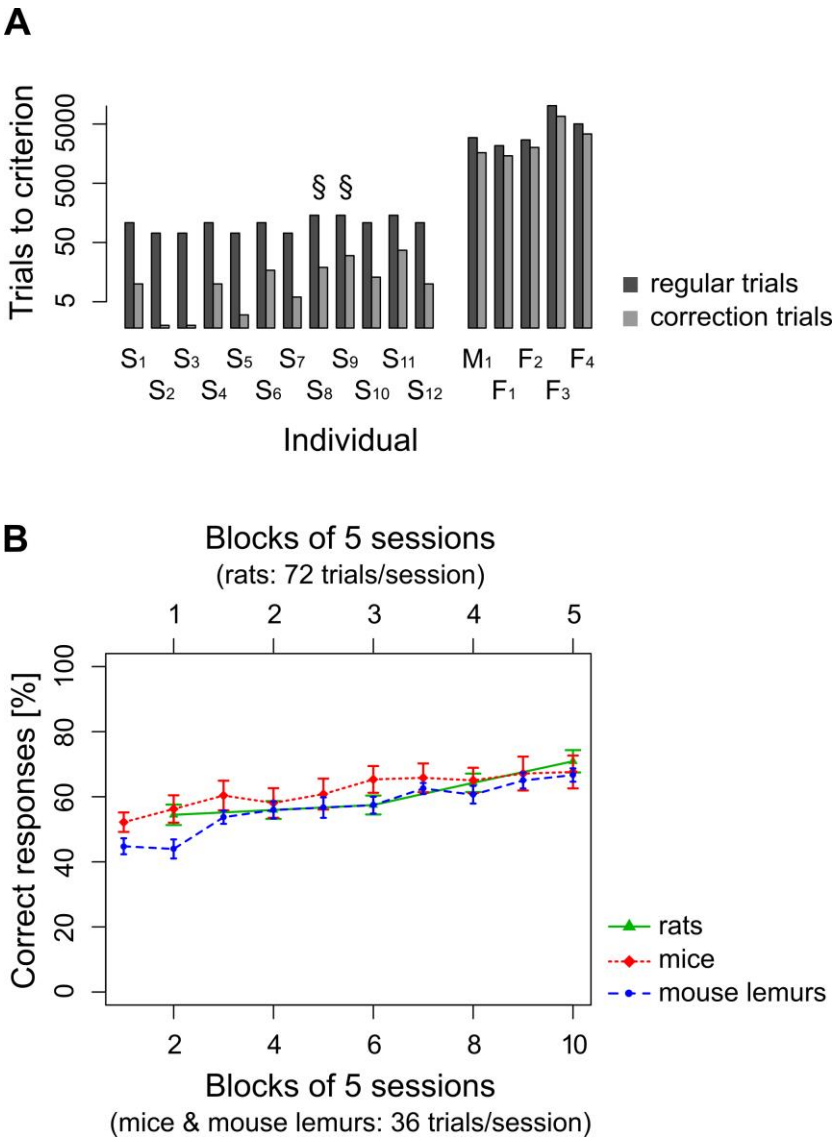

Fig. 3: Cross-species comparisons of dPAL learning performance. A Comparison of the individual number of trials needed to reach the learning criterion between male, human adults (S_1-S_{12}) and mouse lemurs (M_1, F_1-F_4); please note that the ordinate is scaled logarithmically (for a non-logarithmic presentation of the human data see Fig. S5); **§** human subjects that self-reportedly reached the criterion by memorizing all possible stimulus combinations (SC_1-SC_6) instead of finding out the visuo-spatial rule behind the task. **B** Learning performance of the young mouse lemurs as compared to literature values for young, male Lister Hooded rats (Talpos et al., 2009) and young, male C57BL/6 mice (Bartko et al., 2011); Values are presented as group means ±SEM (N_{rats}=7; N_{mice}=7; $N_{mouse\ lemurs}$=3).

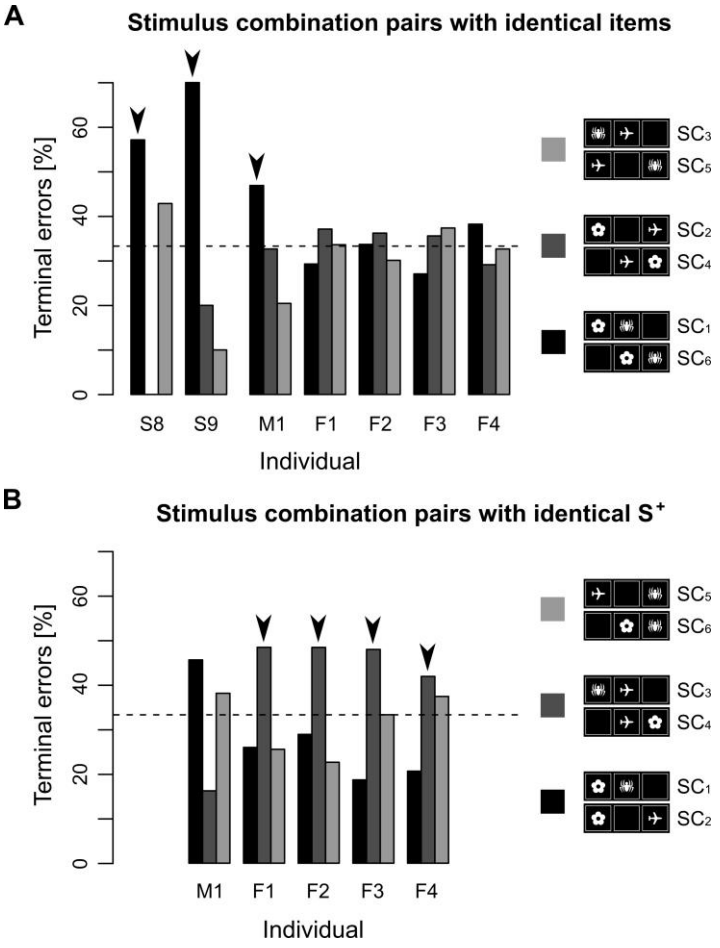

Fig. 4: Individual distribution of terminal errors (last third) compared between humans that solved the task by memorizing all possible stimulus combinations (S_8 [n=7] and S_9 [n=10]) and mouse lemurs (M_1 [n=548], F_1 [n=485], F_2 [n=671], F_3 [n=2251], F_4 [n=1133]). **A** Error distributions separated into stimulus combination pairs with identical items (SC_1/SC_6, SC_2/SC_4, SC_3/SC_5). The two human subjects (S_8 and S_9) showed a clear overrepresentation of terminal errors in SC_1/SC_6 (black arrow heads). In *M. murinus*, a similar pattern with a significant overrepresentation of terminal errors in SC_1/SC_6 (black arrow head) and medium effect size (φc=0.229) was found in M_1; dashed line = chance level (33.3%). **B** Error distributions separated into stimulus combination pairs with identical S^+ (SC_1/SC_2, SC_3/SC_4, SC_5/SC_6). In *M. murinus*, F_1-F_4 showed a significant overrepresentation of terminal errors in SC_3/SC_4 (black arrow heads) with medium effect sizes (φc=0.19 to 0.25); dashed line = chance level (33.3%).