Formatted: Font color: Auto

Two Variations and One Similarity in Memory Functions Deployed by Mice and Humans to Support Foraging

Spencer Talbot, Todor Gerdjikov and Carlo De Lillo

Department of Neuroscience Psychology and Behaviour, University of Leicester

Correspondence should be addressed to:

Carlo De Lillo
Department of Neuroscience, Psychology and Behaviour,
George Davies Centre, Room 3/12
University of Leicester
University Road,
Leicester, LE1 7RH, UK
Tel. +44-0116-229-7193
E-mail cdl2@le.ac.uk

Two variations and one similarity in the memory functions deployed by mice and humans to support foraging

Abstract

Assessing variations in cognitive function between humans and animals is vital for understanding the idiosyncrasies of human cognition and for refining animal models of human brain function and disease. We determined memory functions deployed by mice and humans to support foraging with a search task acting as a test battery. Mice searched for food from the top of poles within an open-arena. Poles were divided into groups based on visual cues and baited according to different schedules. White and black poles were baited in alternate trials. Striped poles were never baited. The requirement of the task was to find all baits in each trial. Mice's foraging efficiency, defined as the number of poles visited before all baits were retrieved, improved with practice. Mice learnt to avoid visiting un-baited poles across trials (Long-term memory) and revisits to poles within each trial (Working memory). Humans tested with a virtual-reality version of the task outperformed mice in foraging efficiency, working memory and exploitation of the temporal pattern of rewards across trials. Moreover, humans, but not mice, reduced the number of possible movement sequences used to search the set of poles. For these measures interspecies differences were maintained throughout three weeks of testing. By contrast, long-term-memory for never-rewarded poles was similar in mice and humans after the first week of testing. These results indicate that human cognitive functions relying upon archaic brain structures may be adequately modelled in mice. Conversely, modelling in mice fluid skills likely to have developed specifically in primates, requires caution.

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

A key issue in psychological theory since the birth of experimental psychology has been the extent to which different species deploy similar mental abilities to humans, perhaps with the exception of language (Bitterman, 1984; McPhail, 1987; Romanes, 1888; 1892). This is critical for the characterisation of what is unique about human cognition and for the understanding of its architecture and evolutionary history (Carlen, 2017; Carruthers, 2013; Passingham, 2008). It has been argued that comparisons between humans and non-human primates, because of their close taxonomic distance and allegedly similar brain organisation, allow the best possible inferences about human cognitive idiosyncrasies and their evolution (Antinucci, 1989; De Lillo, 2012; 2019; Uylings, Groenewegen & Kolb, 2003). Nonetheless, key psychological processes and competences can be observed in both humans and animals which are taxonomically very distant from them such as rodent and bird species (e.g. Davis, 1992; Von Fersen et al., 1991) and can be explained by basic associative mechanisms which pertain to most animals (Frank, Rudy, Levy & O'Reilly, 2005). Many of the experiments highlighting the generality of psychological processes across animal species and humans are carried out on rats and pigeons as animal models. The direct comparison of mice with humans has become particularly important since the development of genetically altered (GA) animal models suitable for genetic targeting neuronal populations and circuits (Carlen, 2017) and which mimic human cognitive dysfunctions (Perlman 2016), especially those due to Alzheimer's disease (McGowan, Eriksen & Hutton, 2006; Zahs & Ashe 2010). Mice are one of the most common GA animal species and are extensively used in behavioural research that aims to translate experimental results obtained with animals onto humans (Perlman 2016). Thus, it has become particularly important to determine the extent of the similarity of cognitive processes in mice and humans.

It has been pointed out that behavioural tests of mice need to tap naturalistic tasks and not many of the commonly used paradigms may be suitable tests (Gerlai and Clayton, 1999).

For example, mice perform poorly compared to rats in the water maze (Gerlay and Clayton, 1999), which is considered the gold standard for the assessment of hippocampal memory functions in rodents (Morris, 1981; Schoenfeld, Schiffelholz, Beyer, Leplow & Foreman, 2017). However, this may not be a true expression of interspecies variations in cognitive factors. Rats, being a species adapted to wetland habitats, find it quite natural to swim. By contrast mice, which are adapted to grassland or forest habitats, can get stressed, freeze and float when immersed in water (Gerlai & Clayton, 1999). This affects the time taken to find a submerged platform, which is the goal of the task. For this reason it has been stressed that the ethology of the species is of pivotal importance for a valid assessment of cognitive skills (Gerlai and Clayton, 1999) and diseases (Perlman 2016). Thus, it is important to test mice in naturalistic conditions that match their ethology. However, it is equally important to devise paradigms that can be used to test humans in conditions that are similar to those used with mice and ideally naturalistic for humans too.

The use of ecologically valid and everyday tasks can be critical for the accurate assessment of cognitive and memory functions in humans (Bailey et al., 2010; Burgess et al., 2006). However, most real-world and present-day tasks that are ecologically valid for humans are not suitable for interspecies comparisons as they require answering questions about daily activities or completing errands in modern day shopping or work environments. Additionally, from an evolutionary perspective it may be important to test humans in conditions and domains that are similar to those that led to the emergence of the functions assessed (Cosmides, 1989; Nairne et al., 2012). The use of Virtual Reality technologies has made it possible to test humans in maze-like tasks that are similar to those used with rodents (De Lillo, Kirby and James, 2014; Redhead and Chan, 2017; Schoenfeld, et al., 2017). Recent studies have devised virtual reality (VR) versions of the water maze for comparing humans

and mice (Schoenfeld et al., 2017) which highlighted similarity in performance between the species but differences in the strategies supporting performance.

For the reasons below, foraging and related large scale search tasks may prove particularly suitable for characterising similarities and variations in the cognitive skills deployed by humans and mice in semi-naturalistic contexts that pertain to both species. Search among multiple targets is considered a ubiquitous process in humans and other species. In fact, search processes identified by behavioural ecologists in a variety of species pertain to humans too (Raichlen, Wood, Gordon, Mabullad, Marlowee, & Pontzerf, 2014) and characterise the way humans search semantic memory as well their external environment (Hills, Todd, Lazer, Redish, Couzin & Cognitive Search Research Group, 2015). Mice spontaneously search large sets of multiple targets placed in open arenas (Crowcroft, 1966; Valsecchi, Bartolomucci, Aversano & Visalberghi, 2000) as do humans in similar search tasks implemented in virtual reality (De Lillo, Kirby & James, 2014). Importantly, as argued below, simulated foraging tasks have the potential to test specific predictions about similarities and variations in cognitive function between mice and human. Similar search processes seem to be shared by humans and a variety of species, including invertebrates (Hills, 2006). However, it is unlikely that search and foraging skills shown by organisms that lack a neocortex require cognitive functions which are analogous to human higher cognitive processes. Yet, the pressures posed by efficient foraging have been considered the trigger for the emergence of large brains and higher cognition in humans and other primate species that share their evolutionary lineage (Aiello & Wheeler, 1995). The foraging hypothesis (Milton, 1993) is one of the main theories of the emergence of large brains in primates (Dunbar and Schultz, 2017) and has recently received renewed support from more extensive neuroanatomical analyses (DeCasien et al., 2017). Specifically, it postulates that efficient foraging for fruit in a forest environment requires the deployment of advanced memory skills.

Fruit in a tropical climate can be a very ephemeral resource. It can ripen and rot in a matter of hours (Milton, 1981). Allegedly, being able to memorise the physical appearance and location of trees of fruiting species would provide an advantage in those environments. Assuming that these skills are mediated by neocortical functions, they would explain the expansion of the frontal cortex that is evident in humans, chimpanzees and other species with a prevalently frugivorous diet (Milton, 1993). Anatomical comparisons of brain and digestive system across frugivorous and non-frugivorous primate species support this notion (Milton, 1993). Recent field studies provide further strong evidence of this to be the case. They show that chimpanzees and other frugivorous species have a remarkable ability to identify the visual appearance of fruiting trees, their location and patterns of fruiting synchrony and alternation in producing ripe fruit (Ban, Boesch & Janmaat, 2014; Janmaat, Ban and Boesch, 2013a, b). Despite the widespread use of laboratory mice in behavioural research, there is surprisingly little information regarding their foraging patterns (Perlman, 2016). Nonetheless their tendency to forage opportunistically and mostly on granary resources that are less ephemeral than fruit and more evenly distributed over large areas has been pointed out (Perlman, 2016; Valsecchi et al., 2000). Therefore, it is possible that the cognitive skills that mice are able to deploy to sustain foraging reflect their specific foraging adaptations and are different from the functions that humans would deploy in the same task.

This study aimed to compare mice and humans in a foraging task and determine if there is a variation in some of the cognitive skills that they deploy in this task, in accordance with the above predictions. In order to do so, for the present study we developed a simple foraging task that is suitable for testing mice in an open arena containing poles, a subset of which are baited, and humans in a VR version of the task. The task affords a simple intuitive measure of foraging efficiency (FE): the number of times poles are visited before all hidden rewards are retrieved in each trial, representing a foraging bout. Critically, this measure of FE can be

decomposed into component measures, each tapping a different memory skill: 1) Long-term memory (LTM) for poles that never yield reward and should be avoided; 2) working memory (WM) for visits of poles already visited within a trial and 3) the tendency to exploit temporal patterns (TP) according to which different subsets of poles are rewarded across trials. These measures of memory function are all subsumed by the measure of foraging efficiency as defined above, rather than being associated with foraging behaviour arbitrarily or by conjecture. As the foraging task affords the simultaneous assessment of multiple memory functions that can be spontaneously deployed to enhance efficiency, it can be used as a mini test battery. Hopefully, such a battery provides the opportunity to overcome difficulties with mapping a taxonomy of human memory functions onto what are often considered equivalent functions in rodents but are not operationally defined in the same way. For example, one important type of LTM function is episodic memory. However, there are doubts concerning whether or not humans and rodents encode episodes in LTM using the same type of temporal information (Roberts et al., 2008) and if the retention of multiple episodes is possible at all in rodents unless olfactory cues are provided as part of the task (Panoz-Brown et al., 2016). This limits the possible use of rodents as suitable models of Alzheimers' disease, which in humans becomes debilitating with the widespread loss of multiple episodes (Panoz-Brown et al., 2016).

Formatted: Font color: Auto

Another type of LTM ostensibly related to foraging is spatial memory. Just like other types of memory, spatial memory is not a unitary construct and is assessed in many different ways in humans (Kessels et al., 2001; Shah et al., 2013). In rodents it is often studied in relation to the mechanisms responsible for learning a location to escape an aversive situation (e.g. in the water-maze) or to find a food reward (e.g. Cheng, 1986). One of the main objectives of spatial memory research has been to determine if global geometric environmental information or general associative mechanisms relying on local cues are used

when learning how to return to a specific goal location (Cheng, K., & Gallistel, 2005; McGregor et al., 2006). Participants are often tested following manipulations of their relative position in relation to proximal or distal cues to see which manipulation affects their spatial searches and from this infer which cues and frames of reference guide learning, sometime from a developmental perspective (e.g. Leplow et al., 2003). Rather than focusing on what type of representation mediates learning of goal positions we aimed to determine similarities and differences in foraging efficiency and its component parts in humans and mice observed in naturalistic foraging conditions simplifying those reported in field studies (e.g. Janmaat, et al., Ban and Boesch, 2013a, b). These conditions afford learning sets of multiple locations (which can often be categorised on the basis of non-spatial visual characteristics such as those defining different plants) and the temporal pattern according to which they yield rewards.

Formatted: Font color: Auto

Formatted: Font color: Auto

Within foraging bouts (here simplified as individual trials) it becomes important to keep track of locations visited. Spatial WM and strategies that may off-load it become important to achieve this. Spatial WM is operationally defined in different ways in the literature. The assessment of spatial WM in human cognitive psychology is based on span tasks requiring the serial recall of observed movement sequences (Baddeley, 2003; De Lillo, 2004). However, the ability to imitate serially ordered movements is unlikely to underpin naturalistic foraging and it is almost impossible to assess it in rodents. Spatial WM measures based on search behaviour in the radial maze, where it is beneficial to keep track of arms visited in each trial (Olton and Samuelson, 1976), is more pertinent within a comparative framework. However, searching in the radial maze may not tap basic WM skills which are similar to those assessed in humans with memory span tasks (Carruthers, 2013). WM as required in naturalistic foraging is likely to be akin to complex WM, which can be defined as the short-term retention of information alongside activation of LTM (Unsworth & Engle. 2006). This would plausibly be necessary to keep track of locations visited in a foraging bout,

while activating LTM for learned information about the environment that can help constrain search (e.g. the visual characteristics of plant species which do not yield fruit and are not worth approaching).

Finally, mapping the ability to detect spatio-temporal patterns in human and animals and establish similarity of function is not a trivial feat either. The encoding of spatio-temporal structure in humans is associated with frontal and parietal activation when measured with span tasks (Bor et al., 2013). In rodents, the detection and deployment of spatio-temporal patterns is assayed in very different contexts such as alternation behaviour in Y and T mazes. A distinction is made between spontaneous alternation patterns, which in mice are mediated by hippocampal functions, and alternation as an adaptive response to alternating reward contingencies. The latter function, is more likely related with frontal functions in rodents (Deacon and Rawlins, 2006; Lalonde, 2002). Alternation between multiple groups of locations rather than between the binary choices afforded by T and Y is more likely to subtend naturalistic foraging and be mediated by fluid skills but it is seldom investigated.

In relation to the above ambiguities in the definition of memory constructs which pertain to humans and mice alike, we considered the task used for this study ideal for identifying similarities and variations in foraging patterns spontaneously displayed by humans and mice tested in similar conditions and, by doing so, highlight analogies and differences in their cognition.

Our prediction was that humans would show a particularly strong tendency to deploy fluid skills such as working memory (including strategies that may be used to offload it) and the ability to detect and exploit temporal patterns of resource availability. Humans may do so in addition to the deployment of LTM skills mediated by brain structures that emerged to sustain navigational systems in early vertebrates and which may be common in several modern mammalian species (Murray, Wise and Graham, 2017). The extent to which mice

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

deploy similar cognitive functions and, as such, can be considered suitable models of human cognitive function will be revealed by their relative performance on measures of these tendencies.

Method

Mice

Subjects. Six C57BL/6J male mice (*Mus Musculus*) were purchased from the Pre-Clinical Research Facility at the University of Leicester (UOL) and were housed in two groups of three animals each, with an inverted 12:12 light-dark cycle. The animals were tested during the dark phase. The animals were aged 18 weeks at the start of the experiment and had been previously exposed to a familiarisation phase (see below) for two weeks. The mice were not food deprived. Diet was removed at 8:30 every morning and was returned when all mice completed their daily testing session. Testing started at 10 am and was completed by 2:30 pm. Individual mice testing order was counterbalanced according to a Latin square design. The UOL Animal Welfare and Ethical Review Body approved the experimental protocols used in this study. Consultation with the Home office confirmed that the study did not meet the threshold for regulation under the Animal (Scientific procedures) Act (ASPA) as it was not considered to involve food deprivation or cause distress to the animals.

Apparatus. A schematic presentation of the apparatus and procedure used with mice is provided in figure 1A. The apparatus used with mice was an open arena measuring 60cm in diameter. It contained a set of plastic structures, each comprising a cylindrical base (5cm tall) surmounted by a pole 7 cm tall (henceforth referred as poles). Thus, the overall structure was 12 cm tall. The poles were designed so that the mouse could not reach the content of the food well at the top of the pole or look into it without stepping on the base and rearing. The arena was surrounded by a thin opaque plastic wall which was 20cm tall. The apparatus was placed

on a table in the centre of a testing room illuminated by a 60w red bulb. A video camera was mounted above the arena. It allowed the experimenter to observe the behaviour of the animals during the experiment from an adjacent room. The poles within the arena were secured to the table using blue-tac. A set of 12 poles was used in the experiment. It comprised 4 white poles, 4 black poles and 4 poles with black and white stripes (Fig. 1A). Stripes were applied to the poles by adding white tape to a black pole. We used very small sugar pellets (Simply Topps mini sugar balls, 1 mm in diameter approx) as rewards.

Figure 1

Procedure: Familiarisation phase. A familiarisation phase took place before the experiment commenced. Mice were handled once a day by the experimenter for a few minutes for a week. From the following week, mice were exposed to the arena, at the beginning without any poles. A handful of sugar pellets was scattered on the floor of the arena and mice were allowed to eat them for 10 min. Then, a grey pole was introduced at random locations and baited with a pellet. The grey colour for the pole used in the familiarisation phase was chosen as it was not used in the experiment to avoid carry-over effects. Other pellets were scattered on the pole's base and around the arena in proximity of the pole. When all mice retrieved the pellet from the top of the pole in each of three consecutive one-trial daily testing sessions, the experiment started.

Procedure: Experiment. A schematic representation of the spatial arrangement of the poles and the baiting of the poles with rewards across trials is presented in Figure 1A.

Mice were tested individually. At the beginning of each trial, each mouse was placed in the centre of the arena and allowed to search the arena and visit the poles. Each trial lasted until the mouse collected all four sugar pellets or when 10 minutes elapsed, whatever

occurred first. It turned out that none of the trials needed to be terminated because the time limit was reached.

Once the trial ended, the mouse was removed from the arena and placed into a temporary holding cage. A surface cleaner (Distel High Level Disinfectant) was applied to the apparatus using absorbing paper to remove any possible odour cues, before the next trial ensued. A visit to a pole was operationally defined as observing the animal step on the base of the pole with at least one paw and rear (see figure 1B). On a daily basis, a sample of the scoring was double checked by a second experimenter from a video recording of the testing session to ensure consistency of scoring.

The baiting of white and black poles alternated across trials. The first trial of a testing session, all four black poles were rewarded, in trial two, all white poles were rewarded, in trial three, black poles were rewarded and so on (See figure 1A). White and black poles were baited in an equal number of trials. Each daily testing session comprised a total of 12 trials. Mice were tested 5 days a week for 3 weeks. Thus, a total of 15 sessions took place. In total, mice undertook 180 trials.

Humans

Participants Eight participants (two males and six females) recruited through the University of Leicester's participant panel took part in the study. Participants had an average age of 27.63 years (SD = 7.69 years), normal or corrected to normal vision and received a reward of £80 on completion of all the 15 testing sessions.

Apparatus and procedure: The apparatus used with humans is depicted in figure 1C. It consisted of a PC running Vizard 4 (WorldWiz Inc) to simulate the environment used to test mice in an immersive virtual environment. Participants saw the environment via either an Oculus rift headset or Nvidia 3D glasses. They controlled their movements in the environment using a PlayStation 4 controller. They used a thumb-stick to control their

movements. When a pole was approached and the participant view point reached a close distance from its base, a written message appeared on the screen with the text "Would you like to visit this pole?" At this point, the participant could operate a trigger if they wished to visit the pole to assess the presence of reward. When a rewarded pole was visited, an icon representing a pair of cherries was displayed (see figure 1D). When a non-rewarded pole or a previously selected rewarded pole was visited, the cherries were not displayed. No additional feedback was given. The virtual environments was carefully designed to mimic the layout of the environment used with mice, including the light source from the top of the room, the shades of the poles on the floor of the arena and as much detail as possible was made similar to the open arena used with mice. The visual cues which could be seen in the virtual environments by looking up onto to the ceiling external to the virtual apparatus were made a similar as possible to those use by the mice. We did so by taking a photograph of the ceiling of the mice testing room from the centre of the arena and reproducing it as part of the virtual display. As the room where mice were tested was unfurnished and with bare walls, the only visual cues were the red light, the casing that held it in place and the video camera on a bracket. The number of poles and patterns of allocation of rewards across trials was the same as that used with mice (see figure 1A). The familiarisation procedure differed from that used with mice. Humans received one single training trial before the testing proper. In the training trial participants were required to visit all poles, arranged as a square matrix within a simulated room. This was to familiarise subjects to the controls and evaluate possible motion sickness caused by using the Oculus rift headset. If participants experienced an uncomfortable amount of nausea, they were given the option of using the 3D glasses instead of the Oculus rift headset. Five participants used the Oculus headset for all sessions of the experiment. Whereas the remaining three used 3D glasses. The experimental procedure was similar to that used for the mice, apart from the following differences. Humans were given

Formatted: Font color: Auto

instructions to "find cherries hidden among the poles in the arena" by approaching and selecting the poles to visit. All other information about the paradigm was withheld. Instead of rearing near the poles, humans visited them in the virtual environment by standing near the pole's base and pressing "R1" on the PlayStation controller. The same number of sessions and trials were used for humans as it was used for mice. As for mice, humans received 12 trials in each daily testing session, 5 days a week, for 3 weeks. The other difference between the procedure used with mice and humans was that mice were returned to their holding cage in between trials, whereas humans at the end of each trial were immediately returned to the starting position and a new trial ensued.

Data analysis

The same variables were used for data analysis in both mice and humans.

An overall measure of Foraging Efficiency (FE) was used as a dependent variable. It was defined as the number of observed visits to a pole in each trial, before all the rewards were retrieved and the trial terminated. Three component measures of FE were then computed, each tapping a different cognitive function. They were: (a) the number of poles revisited in each trial, which was used as a measure of working memory (WM) errors; (b) the total number of visits of striped poles that were never rewarded across trials was used as measure of long-term memory (LTM) errors; and (c) the proportion of trials where a rewarded pole was the first pole visited at the outset of the trial. We considered this measure to be indicative of the extent to which the subject detected the temporal pattern (TP) of alternation of rewards across trials and used it to guide search.

Results

Both mice and humans engaged with the task and were motivated enough to complete each trial. This happened despite the lack of food deprivation in mice and food or monetary rewards specifically awarded for visiting rewarded poles in people. All the trials ended

because all rewards were retrieved and none ended because the time limit was exceeded. Statistical analyses for Foraging Efficiency and all its component measures are reported one by one below.

Foraging Efficiency (FE)

The average number of poles visited by mice and humans before each trial was completed is reported in figure 2.

Figure 2

A 3 (week) x 5 (days) x 3 (trial block) x 2 (species) mixed measures ANOVA was carried out on FE. It revealed main effects of week [$F(2, 24) = 28.40$, $p < .001$, η_p^2=.70]. FE improved significantly between week 1 and week 2 [$t(13) = 4.50$, $p = .005$] but not between week 2 and week 3. There was significant effect of day of testing [$F(4, 48) = 3.80$, $p = .009$, $\eta_p^2 = .24$). A trend analysis showed that this effect was due to a significant linear component in the improvement of FE across days of the week [F(1, 12) = 6.44, p = .026, η_p^2= .35]. There was an effect of trial block [$F(2, 24) = 8.14$, $p < .005$, η_p^2= .470]. FE improved significantly between Block 1 and Block 2 [$t(13) = 3.10$, $p = .013$] but not between trial block 2 and trial block 3. There was a highly significant effect of species [$F(1, 12) = 29.73$, $p < .001$ η_p^2= .95, with humans outperforming mice, overall.

A significant interaction week by species was observed [$F(2, 24) = 4.39$, p =.024, η_p^2= .27]. This interaction can be explained by the fact that mice's FE improved between W1 and W2 [t (5) = 5.40, $p = .003$] as well as between W2 and W3 [$t(5) = 2.51$, $p = .026$]. By contrast human's FE improved only between W1 and W2 [$t(7) = 2.51$, $p = .040$]. A significant interaction week by day [$F(8, 96) = 3.65$, $p = .001$, η_p^2= .23] emerged also. In order to explain

this interaction, we carried out a repeated measures ANOVA for each week to assess the effects of day of testing and the presence of trends of FE across days. They revealed a significant FE improvement across days in week 1 [$F(4, 52)= 5.00$, $p = .002$, $\eta_p^2= .28$]. A trend analysis revealed a significant linear component for W1 [$F(1, 13) = 9.32$, $p = .009$, $\eta_p^2= .42$]. By contrast FE did not significantly improve across days in week 2 and 3. This difference in trends between the first and second week of testing can explain the interaction. None of the other main effects or interactions proved significant. To specify further the lack of species effects, we conducted a Bayesian analysis on LTM, which revealed that our results are 2.57 times more likely to be observed under the null hypothesis predicting no difference between species than under H_1.

Working Memory (WM)

The average number of poles revisited in each trial by mice and humans is depicted in figure 3.

Figure 3

A 3 (week) x 5 (days) x 3 (trial block) x 2 (species) mixed measures ANOVA carried out on WM errors (poles revisited in the same trial) revealed the following effects. There was a highly significant main effects of week [$F(2, 24)= 31.79$, $p < .001$, $\eta_p^2= .73$], as well as a main effects of species [$F(1, 12) = 413.20$, $p < .001$, $\eta_p^2 = .97$. There was also a highly significant interaction week by species [2, 24) = 12.83, $p < .001$, $\eta_p^2 = .517$], with humans making fewer WM errors than mice, overall. Planned comparison carried out to explain this interaction showed that there was a significant reduction of WM errors between W1 and W2 in both mice [$t(5) = 4.70$, $p = .005$] and humans [$t(5) = 2.55$, $p = .038$]. For both species the reduction of WM errors reached a plateau after week one as the difference in WM errors

between W2 and W3 was not significant in either mice or humans. A further independent samples t-test was thus carried out on the difference between the WM errors observed in W1 and W2 in each species to determine if the improvement was larger in one species than the other. The test revealed that this difference was significantly larger for mice than for humans [$t(12) = 4.03$, $p = .002$]. Therefore, this difference between species in the amount of improvement from the first to the second week is likely to explain the interaction.

Long-Term-Memory (LTM)

The average number of never rewarded poles visited by mice and human is reported in figure 4.

Figure 4

A 3 (week) x 5 (days) x 3 (trial block) x 2 (species) mixed measures ANOVA carried out on LTM revealed main effects of week [$F(2, 25) = 20.50$, $p < .001$, $\eta_p^2 = .63$], day of testing [$F(4, 48)$, $p < .001$, $\eta_p^2 = .58$] and trial block [$F(2, 25) = 17.10$, $p < .001$, $\eta_p^2 = .59$]. There was not significant main effect of species on this variable.

Planned comparisons revealed that the frequency of LTM errors only decreased between week 1 and 2 [$t(13) = 3.66$, $p = .003$]. The difference between week 2 and week 3 was not significant, indicating that no further improvements occurred with additional weeks of testing. The number of LTM errors decreased according to a linear component across weeks testing [$F(1, 12) = 25.96$, $p < .001$, $\eta_p^2 = .68$], days [$F(1, 13) = 28.51$ $p < .001$ $\eta_p^2 = .70$] and trial blocks [$F(1, 13) = 7.90$, $p = .001$, $\eta_p^2 = .61$]. No significant effects emerged for species in LTM. There was a significant interaction week by day [$F(8, 96) = 16.56$ $p < .001$, $\eta_p^2 = .58$], as well as a third order interaction week by day by trial block [$F(16, 192) = 2.68$, $p = .001$, $\eta_p^2 = .18$]. In order to interpret this pattern of interactions three separate 5 (days) x 3

(trial block) repeated measures ANOVAs were carried out for each week of testing. Only the ANOVA carried out on week 1 revealed significant main effects of day [$F(4, 52) = 21.02$, $p < .001$, $\eta_p^2 = .62$] and trial block [$F(2, 26) = 6.28$, $p = .006$, $\eta_p^2 = .33$]. No significant main effects emerged for week 2 and 3, suggesting that for both species learning took place exclusively within the first week of testing. Trend analyses showed that in week 1 LTM errors decreased according to a significant linear component across days [$F(1, 13) = 29.63$, $p < .001$, $\eta_p^2 = .70$] and trials blocks [$F(1, 13) = 7.90$, $p = .015$, $\eta_p^2 = .38$]. No significant main effects or interactions emerged from the ANOVAs carried out on the week 2 and 3 of testing.

Detection and use of temporal pattern (TP) of alternation

Formatted: Font color: Auto

The proportions of trials where mice and humans selected a rewarded pole as their first visit to a pole at trial outset is presented in figure 5.

Figure 5

As these proportions were approximately normally distributed for both mice and humans (Shapiro-Wilk $p > .05$ for both species) we analysed the data using t-tests. Humans showed a significantly higher proportion of such trials than mice [$t(12) = 3.89$, $p < .001$].

We then carried out one-samples t-test for each species to assess whether their selections of rewarded poles at trial outset was significantly different from chance. We first tested the test value of .33, as only one third of the poles is rewarded in each trial. Mice selections proved above chance when the test value of .33 was used [$t(5) = 15.79$, $p < .001$]. We ran an additionally one-sample t-test using the more conservative test value of .5. The latter was done considering that both mice and humans rapidly learned to avoid never rewarded poles (as shown by the results obtained for LTM measure) and thus possibly confined their searches to white and black poles only. If this was the case, the probability of

searching a rewarded pole at trial outset would have been .5. Mice were not significantly above chance level when the test value of .5 was used [$t(5) = 1.83$, $p = .127$]. By contrast, humans selected rewarded locations at trial outset above chance level irrespective of whether the test value of .33 [$t(7)= 7.94$, $p < .001$] or the more conservative tests value of .5 was used [$t(5)= 4.98$, $p = .005$].

Interaction of species by component measures of FE.

Finally, to test statistically whether the three different component measures of FE (WM, LTM and TP) were differentially affected by species, we conducted a 3 (component measure) x 2 (species) mixed ANOVA with normalised scores on each measure as the dependent variable. Scores were normalised for each measure to remove scaling effects. Normalised scores for the TP variable were reversed to align them to the other variable that recorded errors rather than correct responses. They are reported in Figure 6.

Figure 6

The ANOVA revealed a significant species x component measure interaction, [$F(2, 24) = 9.75$, $p < .001$, $\eta_p^2 = .45$]. There was also a significant main effect of species [$F(1, 12) = 14.27$, $p = .005$, $\eta_p^2 = .54$] with humans outperforming mice overall. However, the main effect of component measure was not significant [$F(2, 24) = .199$, $p = .821$]. Pairwise comparison confirmed that the interaction species by component measure of FE was explained by the fact that there was a significant interspecies difference in WM [$t(12) = 20.35$, $p < .001$] and TP [$t(12) = 7.62$, $p = .005$] but not in LTM [$t(12) = .489$, $p = .586$]. This latter analysis allows us to confirm that among the different components of foraging efficiency as measured in our task, there are clear cut interspecies differences in WM and TP but not in LTM.

Analysis of sequences

It has been pointed out that self-regulatory processes, occurring when repeatedly practising a search task, can lead to the development of fixed search sequences (De Lillo, Visalberghi & Aversano, 1997; De Lillo, Aversano, Tuci & Visalberghi, 1998; De Lillo, 2012; 2019; Hills et al., 2014). Using one fixed sequence, or a small subset of sequences, among all the paths which can be taken to explore a set, can have the effect of off-loading working memory. This is because familiar paths would help keeping track of poles visited without maintaining them in working memory. In fact, being at any step of a familiar fixed sequence informs the searcher of the locations that have already been explored. As such, fixed sequences can be used as a strategy to minimise the cognitive costs of searching efficiently. In order to determine whether the two species showed a difference in the development of such search strategy, we analysed the number different sequences used by humans and mice to explore the set of poles during the three testing weeks.

Sequences were defined as the serial order in which poles were visited in each trial and analysed separately for trials where black (black trials) or white poles (white trials) were rewarded. Figure 7 reports the number of different sequences observed in black and white trials for each individual across the three weeks of testing. From the figure it can be observed that in humans the number of different sequences performed decreased across weeks of testing and for most participants became very low in the third week. By contrast, in mice the number of different sequences remained high and constant across the three weeks of testing.

Figure 7

Goodness of fit chi X^2 tests were carried out to determine if, among the sequences deployed by each subject, some were used significantly more than others. As required by X^2 analyses (Fisher, 1925; Siegel & Castellan, 1988), all sequences which had an observed

frequency below five were grouped into one category with the expected value weighted accordingly.

The analyses showed each of all eight human participants used either one or a few specific sequences significantly more frequently than expected by chance in both black [$82.87 \leq X^2 \leq 772.87$, $1 \leq df \leq 4$, N = 90, all $p < .001$] and white trials [$89.41 \leq X^2 \leq 870.61$, $1 \leq df \leq 3$, N = 90, all $p > .001$].

By contrast, none of the X^2 tests carried out for each of the six mice proved significant, suggesting that mice did not use any subset of sequences more frequently than others. As human participants restricted the number of sequences used to explore the set of poles, Friedman tests were carried out to determine if this number changed significantly across the three weeks of testing. The tests proved significant for both black [x^2 (2) = 139.33, $p < .001$] and white trials [x^2 (2) = 171.50, $p < .001$]. Figure 7 (a, and c), shows that this was due to a reduction in the number of different sequences across weeks. Pairwise comparisons, carried out with the Wilcoxon signed-rank test, confirmed that a significant reduction in the number of sequences used occurred between week 1 and week 2 in both black [$Z = -8.03$, $p < .001$] and white [$Z = -8.92$, $p < .001$] trials. The number of different sequences used kept decreasing significantly between weeks 2 and 3 for both black [$Z = -5.21$, $p < .001$] and white trials [$Z = -7.85$, $p < .001$]. A visual inspection of Figure 7 (b, and d) shows that the number of different sequences used by mice remained high and constant across the testing weeks.

Discussion

The aim of the present study was to determine variations and similarities in the memory functions spontaneously deployed by mice and humans to support foraging efficiency. We used a task deemed to exploit the natural exploratory tendency of mice and that could be

implemented with humans using immersive virtual reality. The task proved effective as a mini battery of memory functions for cross species comparisons. Both mice and humans readily engaged with the task, since they completed each trial, without interruptions or need to terminate it because of exceeded time limits.

Foraging efficiency improved with task practice in both species as shown by significant reductions in the number of poles visited to retrieve the four rewards across trials and across daily testing sessions. This is remarkable considering that there was little incentive for the subjects to search efficiently in this task. Mice improved their foraging efficiency without prolonged food deprivation. Mice diet, otherwise present ad libitum, was only removed 2 to 3 hours before testing and put back in place immediately afterwards. Humans' foraging efficiency improved in absence of any externally imposed incentives apart from the shortening of trial duration, which implicitly derived from searching efficiently.

Overall humans foraged more efficiently than mice. However, of particular interest for comparative purposes was the pattern of skills deployed by the two species to increase efficiency. In fact, a critical characteristic of the procedure adopted here is that it enabled foraging efficiency to be partitioned into several component measures, each tapping a separate cognitive function. This made it possible to determine which specific functions contributed to the observed differences in foraging efficiency in the two species. Interspecies differences were explained by a stronger tendency in humans to avoid WM errors and to detect and use the TP of alternation of rewarded poles across trials. In fact, measures of these two abilities significantly differed across species. We will focus on the discussion of interspecies difference in WM first.

Despite being extensively studied in humans, WM is still not well characterised within a comparative context. Assumptions regarding variations in WM between humans and animals range from the number of items that can be held in WM by different species to the

complete denial of any WM ability in animals (see Carruthers 2013, for a review). For this reason, Carruthers (2013) pointed out that direct comparisons of WM in animals and humans are now desperately needed. Some attempts have been made at comparing humans and nonhuman primates (e.g. Fagot & De Lillo, 2011). However, direct comparisons of humans and mice are uncommon. The present study attempted to provide some information to help filling in this gap.

The term WM was first used in radial maze animal studies and defined as the number of re-entries of arms visited within the same trial (Olton and Samuelson, 1976). This is an operational definition similar to the one adopted here (revisits to visited poles). However, for the reasons below we believe that WM as measured in our task may be a different construct than what is measured in the radial maze. In rodents lesion of the hippocampus leads to an increase of both WM errors and LTM (often referred to as reference memory) in the radial maze. This functional overlap may result from the fact that in the radial maze WM may in fact rely in part on LTM (Carruthers 2013).

Formatted: Font color: Auto

In humans and other primates WM is typically associated with frontal rather than temporal functions (Goldman-Rakic & Leung, 2002). In humans WM correlates with fluid skills and general intelligence. However, this is the case only when complex WM tasks are used to measure it. Complex WM tasks require human participants either to perform a secondary operation while maintaining items in WM or to attend simultaneously to multiple types of information (Engle 2010; Unsworth, and Engle, 2006)). A similar relationship between complex WM and general intelligence has been observed in mice. A general intelligence factor has been identified in mice, which explains a large portion of variance across a range of learning tasks (Matzel et al., 2003). Similarly to what happens in humans, this factor correlates exclusively with complex WM tasks, such as those where mice had to retain information about competing memories and avoid interference between them (Kolata,

Formatted: Font color: Auto

Light, Townsend, Hale, Grossman & Matzel, 2005). It is conceivable that WM in our task is more akin to complex than simple WM. We suggest this because both species showed a significant reduction in WM errors between week 1 and week 2, while, at the same time, they showed a significant reduction in LTM errors. The reduction of errors on both measures was then maintained in the third week of testing. This indicates that subjects kept track of the locations visited within a trial (WM) while, at the same time, they attended to visual characteristics (stored on LTM) of poles that had to be avoided as they never yielded rewards. Holding in mind one type of information while tracking visited spatial locations could classify as a complex WM task. Moreover, humans (at least) clearly did so while also monitoring the temporal pattern (TP) of reward availability across trials, a third type of information.

Mice did not monitor TP to the same extent of humans in our task, as shown by the significant difference in the two species on this measure. In primates the temporal organisation of behaviour is mediated by frontal functions (Carlen, 2017). Moreover, the strategic use of predictable temporal patterns (TP) of resource availability in our task could be considered an instance of a more general ability to pick-up patterns in other task domains (see De Lillo, 2012, 2019 for reviews). Such type of ability is also associated with frontal functions in humans and with activity of the dorsolateral pre-frontal cortex (DLPFC) in particular (Bor, Duncan and Owen, 2003).

Thus, the clear interspecies variations that we observed in WM and TP may reflect differences in higher level cognitive function that in primates are mediated by frontal activity. The extent to which the prefrontal cortex of rodents can be considered analogous to that of primates both in neuroanatomical and functional terms is still controversial (Uylings et al., 2003). However, it has been pointed out on the basis of several types of information that the DLPFC in particular seems to be a specialisation of the primate brain (Carlen, 2017). Thus, it

is possible that mice lack some abilities possessed by humans and other primates because of this.

The extent to which TP, as assessed here, is akin to alternation behaviour in other tasks or not (see Lalonde 2002) requires some comments related to the arguments outlined above. Mice generally are capable of detecting alternation of reward patterns in Y and T mazes and show spontaneous alternation in exploratory behaviour in absence of any reward (e.g. Deacon and Rawlins, 2006). The spontaneous tendency to alternate in mice is normally impaired by hippocampal rather than frontal lesions (see Lalonde 2002 for a review). By contrast, learning temporal alternation of rewards is compromised by frontal lesions in rodents as well as in monkeys (Lalonde 2002). Our task is more akin to the latter and thus likely to be mediated by frontal functions that may be particularly developed in humans. In addition to this, there are clear difference between the foraging task adopted here and Y and T mazes (and occasionally the water maze, e.g. Means, Holsten, Long & High, 1996) which are normally used to test alternation in rodents.

The first difference is that our task features a search among multiple locations subdivided on the basis of visual features. In most maze studies, alternation is between binary alternative locations (e.g. between the left arm of right arm of a T maze or two possible locations of a platform in the water maze). By contrast, to exploit alternation in our task, subjects need to either: 1) consider simultaneously the alternation of multiple individual spatial locations; or 2) focus on the colour of the poles, group them accordingly (e.g. all black poles) and then track the alternation of each colour group.

Formatted: Font color: Auto

The second difference is that in most maze studies, alternation occurs spontaneously, or, in the case of rewarded alternation, the subject needs to track the alternation without much other processing occurring at the same time. By contrast, in our task TP is only one of several sources of information that can be beneficial if detected and monitored (e.g. colour of never

rewarded and location of previously visited poles). The additional (and possibly more similar to naturalistic foraging) challenges posed by our task can explain why mice do not alternate in this task despite their propensity to do so in maze studies.

An additional difference that we observed was in the number of search paths used by the two species. Humans only deployed a subset of all the possible sequences that could be used to explore the set of poles. The particular search trajectories used were different for the different participants and decreased in the course of task practice across the three weeks of testing. This suggests that the narrowing down of the number of sequences used derived from self-regulation processes, and was not caused by biases which were already present at the outset of testing. In contrast with humans, mice did not show a tendency to reduce the number of sequential patterns used to explore the set of poles. This finding is consistent with results that we report elsewhere (De Lillo, 2012; De Lillo; 2019; De Lillo, Kirby and James, 2014), which show that in other search tasks the ability to self-regulate and develop fixed search sequences is a peculiar characteristic of primates (and humans in particular) but not observed in mice (Valsecchi et al., 2000).

In contrast with their clear-cut differences in WM and TP and extent of reduction of sequences used in search, mice and humans showed a similar level of performance LTM in our task. In fact, an interaction emerged between species and the standardised scores of the three component measures of foraging efficiency.

One possible explanation for the similarity in LTM functions in mice and humans, but not in WM and TP could be traced in the divergence of brain evolution in the two species. In the over 90 million years that separate mice and humans from a common ancestor, the most significant structural, molecular and genetic differences that affected brain development and organisation in their respective lineages regard the prefrontal cortex (Rakic 2009; Geschwind and Rakic 2013; Carlen, 2017).

By contrast, the hippocampus and para-hippocampal area may be homologous in all mammals, including humans and mice (Allen & Fortin, 2013; Carruthers, 2013). Functions of these more archaic hippocampal structures could explain the similarity of performance in LTM in humans and mice.

Although they do not support them directly, from an evolutionary standpoint, the above considerations would be consistent with hypotheses concerning the relationship between diet, foraging requirements and the emergence of higher cognitive skills in humans as in other primates (Milton 1981, 1993). Supported by anatomical analyses (DeCasien, 2017), these approaches speculate that efficient fruit foraging in these environments can be facilitated by an ability to identify which tree species yield fruit and which do not. Then, among fruiting trees it is important to determine which trees yield ripe fruit at particular points of time. This happens according to cyclical patterns depending seasons and time of the day. Trees of particular species can be identified on the basis of their physical appearance and location. Observational field studies (Janmaat et al., 2013a, b), have shown remarkable abilities in chimpanzees and other species to process information regarding the botanical characteristic of tree species in their habitat and exploit the fruiting patterns of species in there. These aspects of foraging are mimicked, albeit in a much simplified way, in our task by the availability of rewards in subgroups of poles that can be grouped by visual cues. Humans share the same evolutionary lineage of chimpanzees and other primate species with a largely frugivorous diet (Milton, 1981, 1993). As humans we would show the vestiges of abilities that emerged in recent common ancestors with other frugivorous primate species (White, Lovejoy, Asfaw, Carlson & Suwa, 2009). Such vestiges would include efficient working memory and the ability to detect cyclical patterns of resource availability as well as LTM for locations. All skills deployed to a larger extent by humans in our task.

In contrast to what happens with frugivorous primates, it has been pointed out that mice are mainly opportunistic feeders, with a mainly granivorous diet in the wild or a wide variety of sustenance within a domestic setting (Valsecchi et al., 2001). This often makes food supplies unpredictable or available in fluctuating quantities all year round, unlike fruit. Studies in semi natural settings suggest that mice visit all or most of the available food sources during their foraging bouts rather than becoming selective in their searches (Crowcroft, 1966).

LTM as measured here, could be the vestige of navigational memory systems, centred on the hippocampal complex of early vertebrates, and which are still present in modern rodents, monkeys and humans (see Murray et al., 2017).

The above account would fit the pattern of interspecies difference and similarities observed in this study. Of course it is also possible that variations in the task we used could highlight more similarities between mice and humans than those that we observed. Abilities that seemed to pertain mostly to primates, such as the emergence of learning sets in visual object discrimination abilities (Harlow, 1949), have subsequently been observed in mice when olfactory versions of the original task have been used (Schellinck, Forestell, and LoLordo, 2001). It is possible that the same would apply to the present task. Thus, the extent to which mice may be more likely to spontaneously deploy working memory and pattern detection skills when olfactory cues are used to distinguish between poles rather than visual cues remains to be determined. It is also possible that making additional cues and landmarks available within and outside the arena may help subjects memorise items better and facilitate working memory.

A practical implication of our findings is that human LTM functions, like those assessed by this task, may be those that are best modelled in mice. Mice are one of the most widely used genetically altered (GA) animal model of human neuro-psychiatric conditions

Formatted: Font color: Auto

Formatted: Font color: Auto

(Carlen, 2017), of Alzheimer's disease in particular (Gerlai and Clayton, 1999). However, doubts have been cast on the plausibility of mice as models of human cognition (Carlen, 2017). Our results suggest that scepticism may be warranted in respect to modelling with mice human functions related to working memory and the ability to detect patterns in time and plan search behaviour accordingly.

However, LTM functions as deployed in foraging tasks could be quite similar in the two species. This would make mice suitable subjects in studies aimed at testing possible interventions for the amelioration of LTM in conditions that cause its deterioration in humans.

Translational research with mice models should work in both directions with information obtained with mice used to guide human studies and, vice versa using results obtained with humans to interpret those of studies with mice (Zahs & Ashe, 2010). Determining the range of behavioural processes pertaining to both species is essential for this in relation to psychological research. This study made a contribution to this goal by highlighting two variations and a similarity in the cognitive underpinnings of foraging in mice and people.

As a final comment, we would say that it would be useful for future research to evaluate how the variables assessed in this study map onto those derived from human psychometric scales and other established paradigms for assessing memory functions in mice. Notably, these could include simple and complex span tasks in human and mice as well as measures derived from more conventional maze tasks. The ability to detect patterns of alternation may be associated with tests of frontal and executive functions, such as the Wisconsin card sorting test (Eling, Derck, & Maes, 2008).

Formatted: Font color: Auto

It would be also useful to extend the study to human clinical populations and genetically-altered mice models of human neuropsychiatric disease. This would help

determine if assaying cognitive functions when they are related to each other in a consequential and non-arbitrary way, and spontaneously deployed in a naturalistic behaviour, could become the gold-standard for the pre-clinical study of human disease using appropriate animal models.

References

Aiello, L.C. & P. Wheeler, P. (1995). The expensive issue hypothesis: the brain and the digestive system in human evolution. *Current Anthropology*, 36(2), 199-221.

Allen T, Fortin N (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*, 110:10379–10386.

Anderson, S.W., Damasio, H. R. Dallas Jones, H.R. & Tranel, D. (1991) Wisconsin Card Sorting Test Performance as a Measure of Frontal Lobe Damage, *Journal of Clinical and Experimental Neuropsychology*, 13(6): 909-922.

Antinucci, F. (Ed.). (1989). *Cognitive structure and development in nonhuman primates*. Howe, Lawrence Erlbaum Associates.

Baddeley, A.D. (2003). Working memory: looking back and looking forward. *Nature review neuroscience*, 4: 829-839.

Ban, S.D., Boesch, C. & Janmaat, K.R.L. (2014) Taı̈ chimpanzees anticipate revisiting high-valued fruit trees from further distances. *Animal Cognition*, 17, 1353-1364, DOI 10.1007/s10071-014-0771-y.

Bailey, P.E., Henry, J.D., Rendell, P.G., Phillips, L.H. & Kliegel, M. (2010). Dismantling the "age–prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *Quarterly Journal of Experimental Psychology*, 63: 646-652.

Bitterman, M. E. (1984) Learning in man and other animals. In: Perspectives in psychological experimentation: Towards the year 2000, (p.59) V. Sams, V. & A. Parducci, (Eds). Hillsdale, New Jersey: Lawrence Erlbaum.

Bor, D., Duncan, J., & Owen, A.M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand, *Neuron,* 37(2), 361-367.

Burgess, P., Alderman, N., Forbes, C., Costello, A., Coates, L. M.A., Dawson, D.R., Anderson, N.D., Gilbert, S.J., Dumontheil, I. & Channon, S. (2006). The case for the development and

use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, 12, 194–209.

Carlen, M. (2017). What constitutes the prefrontal cortex? *Science*, 358, 478-482.

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 110: 10371–10378.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, 23(2), 149–178

Cheng, K., & Gallistel, C. R. (2005). Shape parameters explain data fromspatial transformations: Comment on Pearce et al. (2004) and Tommasi and Polli (2004). *Journal of Experimental Psychology: Animal Behavior Processes*, 31,254–259

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how human reason? Studies from the Wason selection task. *Cognition*, 31, 87-276.

Crowcroft, P. (1966). *Mice all over*. London: Fowlis.

Davis, H. (1992). Transitive inference in the rat (*Rattus norvegicus*). *Journal of Comparative Psychology*, 106(4): 342-349.

Deacon, R. & Rawlins, J. (2006). T-maze alternation in the rodent. *Nature Protocols* 1, 7–12.

DeCasien, A.R., Williams, S.C. & Higham, J.P. (2017). Primate brain is predicted by diet but not sociality. *Nature Ecology and Evolution*, 1, 0112.

De Lillo, 2004. Imposing structure on a Corsi-type task: evidence for hierarchical organisation based on spatial proximity in serial spatial memory. *Brain and Cognition*, 55(3): 415-426.

De Lillo, C. (2012). The use of spatial structure in working memory: a comparative standpoint. In: D. McFarland, K. Stenning & M. McGonigle-Chalmers (Eds.), *The Complex Mind: an Interdisciplinary Approach* (pp. 38-54). Chippenham & Eastbourne: Palgrave-MacMillan,

De Lillo C. (2019) Variations in the Beneficial Effects of Spatial Structure and Serial Organisation on Working Memory Span in Humans and Other Species. In: Hodgson T. (Eds) *Processes of*

Visuospatial Attention and Working Memory. Current Topics in Behavioral Neurosciences, vol 41. (pp. 59-83). Cham: Springer.

De Lillo, C., Aversano, M., Tuci, E. & Visalberghi, E., (1998) Spatial constraints and regulatory functions in monkey's (*Cebus apella*) search. *Journal of Comparative Psychology*, 112: 353-362.

De Lillo, C., Kirby, M. & James, F.C. (2014) Spatial Working Memory in Immersive Virtual Reality Foraging: Path Organization, Traveling Distance and Search Efficiency in Humans (Homo sapiens). *American Journal of Primatology*, 76, 436-446.

De Lillo, C., Visalberghi, E. & Aversano, M. (1997) The Organisation of Exhaustive Searches in a "Patchy" Space by Capuchin Monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 111: 82-90.Dunbar, R. I. M., & Shultz, S. (2017). Why are there so many explanations for primate brain evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1727), 1-14

Engle, R. (2010) Role of working-memory capacity in cognitive control. *Current Anthropology*, 51:S17–S26.

Fagot, J. & De Lillo, C. (2011). A comparative study of working memory: immediate serial recall in baboons (*Papio papio*) and humans. *Neuropsychologia*. 49, 3870-3880.Eling, P., Derckx, K., & Maes, R.. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, 67(3), 247–253.

Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh: Oliver & Boyd.

Frank, M.J., Rudy, J.W., Levy, W.B., & O'Reilly, R.C. (2005). When logic fails: Implicit transitive inference in humans. *Memory and Cognition*, 33, 742–750.

Gerlai, R., & Clayton, N. S. (1999). Analysing hippocampal function in transgenic mice: an ethological perspective. *Trends in neurosciences*, 22(2), 47-51.

Geschwind, D. H., & Rakic, P. (2013). Cortical evolution: judge the brain by its cover. *Neuron*, 80(3), 633-647.

Goldman-Rakic, P. & Leung, H.C. (2002). Functional architecture of the dorsolateral prefrontal cortex in monkeys and humans. In Stuss, D.T, Knight, R.T. (Eds). *Principles' of frontal lobe function* (pp. 85-95). New York: Oxford University Press.

Harlow, H.F. (1949). The formation of learning sets. *Psychological Review*, 56, 51–65.

Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, 30, 3–41.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & Cognitive Search Research Group (2015). Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences*, 19(1), 46–54.

Janmaat, K.R.L., Ban, S.D. & Boesch, C. (2013a). Taï chimpanzees use botanical skills to discover fruit: what we can learn from their mistakes. *Animal Cognition*, DOI 10.1007/s10071-013-0617-z.

Janmaat, K.R.L., Ban, S .D., Boesch, C. (2013b) Chimpanzees use long-term spatial memory to monitor large fruit trees and remember feeding experiences across seasons. *Animal Behaviour*, 86, 1183-1205

Kessels, R. P. C., De Haan, E. H. F., Kappelle, L. J., & Postma, A.. (2001). Varieties of human spatial memory: a meta-analysis on the effects of hippocampal lesions. *Brain Research Reviews*, *35*(3), 295–303

Kolata, S., Light, K., Townsend, D.A., Hale, G., Grossman, H.C. & Matzel, L.D., (2005) Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory* 84(3), 241–246.

Lalonde, R (2002). The Neurobiological basis of spontaneous alternation. *Neuroscience and Biobehavioral Reviews*, 26, 91-104.

Leplow, B., Lehnung, M., Pohl, J., Herzog, A., Ferstl, R., & Mehdorn, M.. (2003). Navigational place learning in children and young adults as assessed with a standardized locomotor search task. *British Journal of Psychology*, *94*(3), 299–317.

Matzel, L.D., Han, Y.R., Grossman, H.C., Karnik, M., Patel, D., Scott, N. Specht, S.M. & Gandhi, C.C. (2003). Individual differences in the expression of a "general" learning ability in mice. *Journal of Neuroscience*, 23(16), 6423–6433.

Macphail, E.M. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences*. 10(4): 645-656.

McGregor, A., Jones, P. M., Good, M. A., & Pearce, J. M. (2006). Further evidence that rats rely on local rather than global spatial information to locate a hidden goal: Reply to Cheng and Gallistel (2005). *Journal of Experimental Psychology: Animal Behavior Processes*, 32(3), 314–321.

McGowan, E., Eriksen, J., Hutton, M. (2006). A decade of modelling Alzheimer's disease in transgenic mice. *Trends in Genetics*, 22(5), 281-288.

Means, L.W., Holsten, R.D., Long, M., High, K.M. (1996). Scopolamine- and morphine-induced deficits in water maze alternation: failure to attenuate with glucose. *Neurobiology of learning and memory*, 66, 167–175

Milton, K. (1981). Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. *American Anthropologist*, 83(3), 534-548.

Milton, K. (1993). Diet and primate evolution. *Scientific American*. 269: 86-93.

Morris, R. G. M. (1981). Spatial localisation does not depend on the presence of local cues. *Learning and Motivation,* 12:239-260.

Murray, E.A., Wise, S.P., & Graham, K.S. (2017). *The evolution of memory systems*. Oxford: Oxford University Press.

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., & Blunt, J. R. (2012). Adaptive memory: Enhanced location memory after survival processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 495-501.

Olton, D.S., Samuelson, R.J. (1976). Remembrance of places passed: spatial memory in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 2, 97–116.

Panoz-Brown, D. et al. (2016). Rats Remember Items in Context Using Episodic Memory. *Current Biology*, 26, 2821-2826.

Passingham, R. (2008). *What is special about the human brain?* Oxford: Oxford University Press.

Perlman, R. L. (2016). Mouse models of human disease: An evolutionary perspective. *Evolution, medicine, and public health*, 1, 170–176.

Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nature Reviews Neuroscience*, *10*(10), 724.

Redhead, E. & Chan, W. (2017). Conditioned Inhibition in the spatial domain in humans and rats. *Learning and Motivation*, 59, 27-37.

Raichlen, D.A., Wood, B.M., Gordon, A.D., Mabullad, A.Z.P., Marlowee, F.W., & Pontzerf, H. (2014). Evidence of Lévy walk foraging patterns in human hunter–gatherer. *Proceedings of the National Academy of Sciences of the United States of America*, 11(2), 728–733.

Roberts, W.A, Feeney, M.C., McPherson, K., Petter, M., McMillan, N. & Musolino, E. (2008). Episodic-Like Memory in Rats: is it based on when or how long ago? *Science*, 320(4): 113-115.

Romanes, G. J. (1888). *Mental evolution in man*. London: Kegan Paul, Trench and Co.

Romanes, G.J. (1892). *Animal intelligence*. New York: D. Appleton and co.

Schellinck, H.M., Forestell, C.A., LoLordo, V.M. (2001). A Simple and Reliable Test of Olfactory Learning and Memory in Mice, *Chemical Senses*, 26(6), 663–672.

Schoenfeld, R., Schiffelholz, T., Beyer, C., Leplow, B., & Foreman, N. (2017). Variants of the Morris water maze task to comparatively assess human and rodent place navigation. *Neurobiology of learning and memory*, 139, 117-127.

Shah, D.S., Prados, J., Gamble, J. De Lillo, C. & Gibson, C.L. (2013). Sex differences in spatial memory using serial and search tasks. *Behavioural Brain Research*, 257: 90-99.

Sharpe, Donald (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, 20(8): 1-10. Available online: http://pareonline.net/getvn.asp?v=20&n=8

Siegel, S. & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences*. New York; McGraw-Hill.

Slotnick, B.M., & Katz, H.M. (1974). Olfactory Learning-set formation in rats. *Science* 30: 796-798.

Springer, M.S., Murphy, W.J. (2007). Mammalian evolution and biomedicine: new views from phylogeny. *Biological Reviews of the Cambridge Philosophical Society*, 82: 375–92.

Unsworth, N. & Engle, R.W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language* 54(1): 68-80.

Uylings, H.B.M., Henk J., Groenewegen, H.J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research*, 146: 3–17.

Valsecchi, P., Bartolomucci, A., Aversano, M., & Visalberghi, E. (2000). Learning to cope with two different food distributions: The performance of house mice (*Mus musculus domesticus*). *Journal of Comparative Psychology*, 114(3): 272-280

Von Fersen, C.D.; Wynne, D.L., Delius, J, & Staddon, J.E.R. (1991) Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*; 17 (3): 334-341.

White, T.D., Lovejoy, C.O., Asfaw, B., Carlson, J.P., & Suwa, G. (2009). Ardipithecus ramidus and the Paleobiology of Early Hominids. *Science*, 326:75-86

Formatted: Font color: Auto

Formatted: Font color: Auto

Zahs, K.R., & Ashe, K.H. (2010) 'Too much good news' – are Alzheimer mouse models trying to tell us how to prevent, not cure, Alzheimer's disease? *Trends in Neurosciences* 33: 381–389.

Formatted: Font color: Auto

Figure captions

Figure 1

A schematic representation of *apparati* and procedure used with mice and humans. A. Diagram of the arrangement of poles in the arena used with mice and the virtual reality arena used with poles. In figure 1A, is also shows the pattern in which poles of different colours were baited across the 180 trials featured in the testing phase. Plus and minus signs indicate the presence or absence of reward in that particular trial. Black and white poles were rewarded in alternate trials. Striped poles were never rewarded (see text for a more detailed explanation). B. The set-up used to tests humans. C. schematic representation of a pole selection by a mouse. The pole was considered selected if the mouse stood on the based on the pole and reared. D. Screen shot showing the results of the selection of a rewarded pole by a human in VR. Humans used the trigger of handheld control to select a pole when in proximity of it (see text for a detailed explanation).

Figure 2

Foraging efficiency.

Average number of poles visited by mice (white circles) and humans (black circles) before each trial was completed. The X axis reports days of testing. Each daily testing session was divided into 3 blocks of 4 trials each. Results for the first, second and third week of testing are reported in the top, middle and bottom graph, respectively. Error bars represent 1 SE above and below the mean.

Figure 3

Working memory (WM) errors

Average number of revisits of poles already visited in the same trial by mice (white circles) and humans (black circles). The X axis reports days of testing. Each daily testing session was divided into 3 blocks of 4 trials each. Results for the first, second and third week of testing are reported in the top, middle and bottom graph, respectively. Error bars represent 1 SE above and below the mean.

Formatted: Font color: Auto

Figure 4

Long-term memory (LTM) errors

Average number of visits per trial of poles that are never rewarded (striped poles) by mice (white circles) and humans (black circles). The X axis reports days of testing. Each daily testing session was divided into 3 blocks of 4 trials each. Results for the first, second and third week of testing are reported in the top, middle and bottom graph, respectively. Error bars represent 1 SE above and below the mean.

Formatted: Font color: Auto

Figure 5

Use of Temporal Pattern of alternation of rewarded poles

Proportion of trials in which mice (white circles) and humans (black circles) selected a rewarded pole as their first choice at trial outset. Dotted lines represent .33 and .50 proportions used as test values in one sample t tests. ** = $p < .001$ in the comparison between species. See text for explanation. Error bars represent 1 SE above and below the mean.

Figure 6

Normalised scores for the tree component measures of foraging efficiency: Working memory (WM), use of temporal pattern (TP) and long term memory (LTM). Error bars represent 1 SE

above and below the mean. Values for mice are depicted as white circles. Values for humans are depicted as black circles. Error bars represent 1 SE above and below the mean.

Figure 7

Number of different search sequences observed in each of the three weeks of testing for humans (left: a, and c) and mice (right: b, and d) in trials where black (top: a, and b) or white (bottom: c, and d) poles were rewarded. Error bars represent 1 SE above and below the mean.

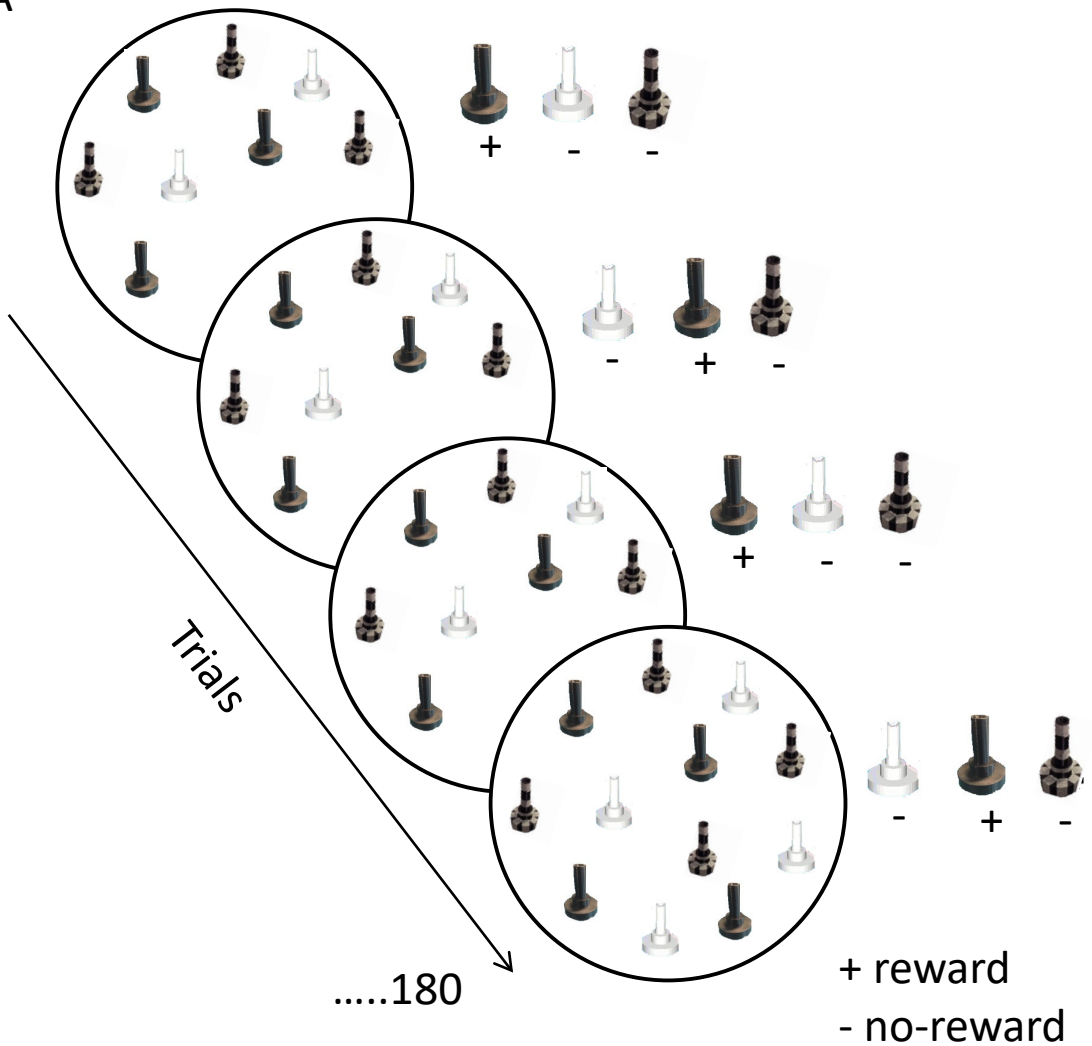
Acknowledgments

We are extremely grateful to Richard Kirkden for developing the software used to generate the virtual environment used with humans; all staff at the Pre-Clinical Research Facility at the University of Leicester for their assistance with the maintenance of mice; Gonzalo Urcelay for advice on the interpretation of Bayes factors; and an anonymous reviewer, who suggested to analyse individual behaviour in relation to the use of fixed sequences in search.

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a small grant by the Experimental Psychology Society to Carlo De Lillo and Todor Gerdjikov.

Fig 1

A

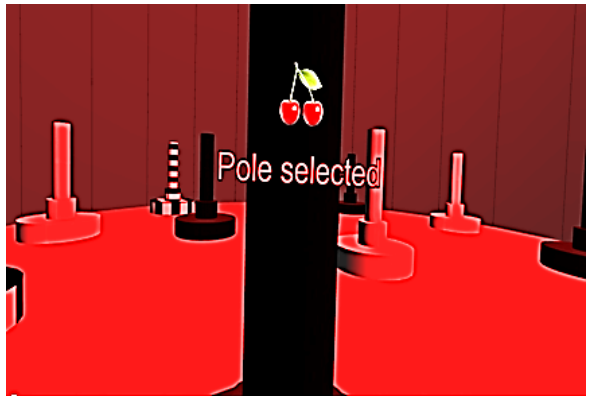

+ - -
- + -
+ - -
- + -
Trials
.....180
+ reward
- no-reward

B

C


Oculus
Pole selected

D


Pole selected

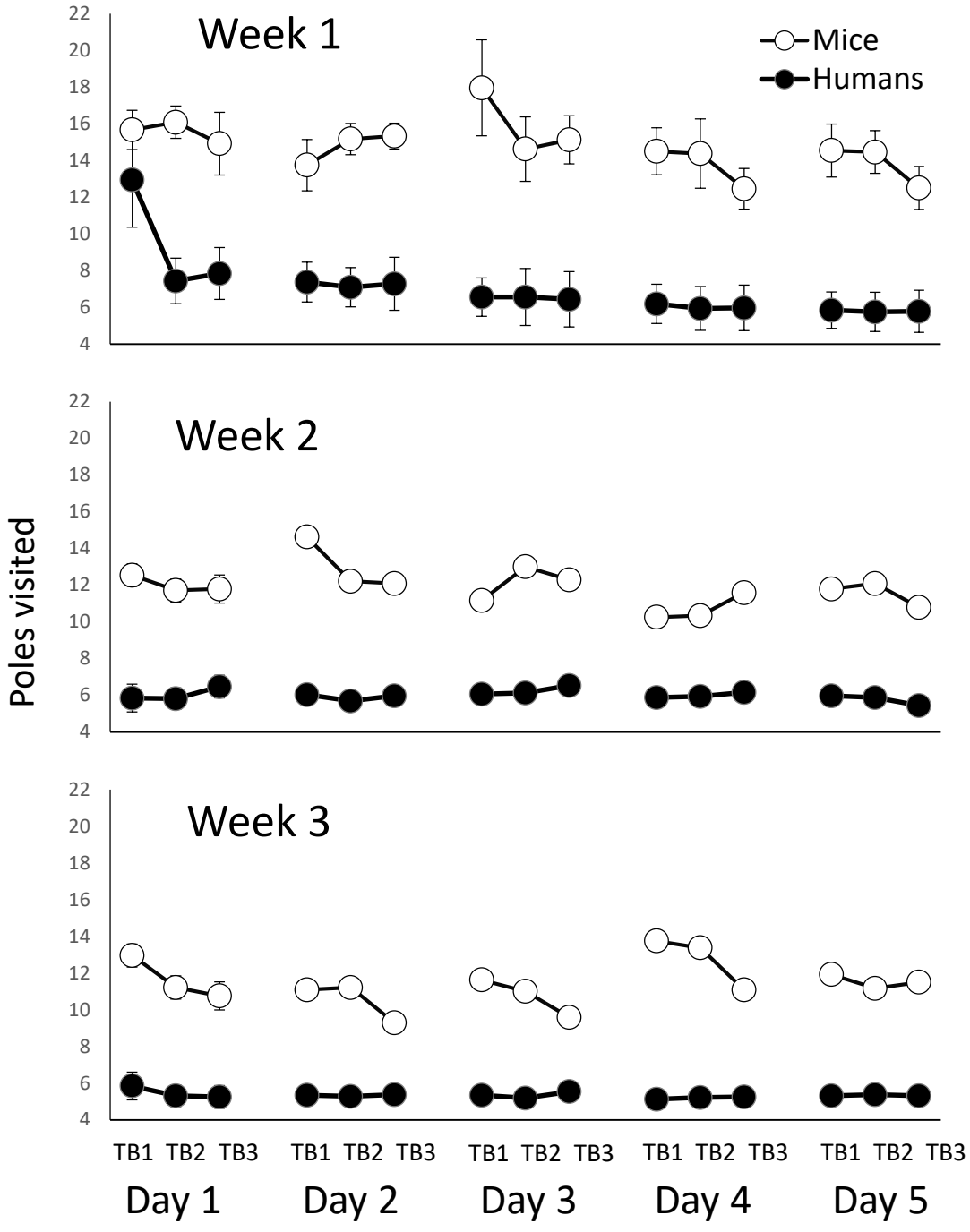

Week 1
Mice
Humans
Week 2
Week 3
Poles visited
22
20
18
16
14
12
10
8
6
4
TB1 TB2 TB3
Day 1
Day 2
Day 3
Day 4
Day 5

Fig 2

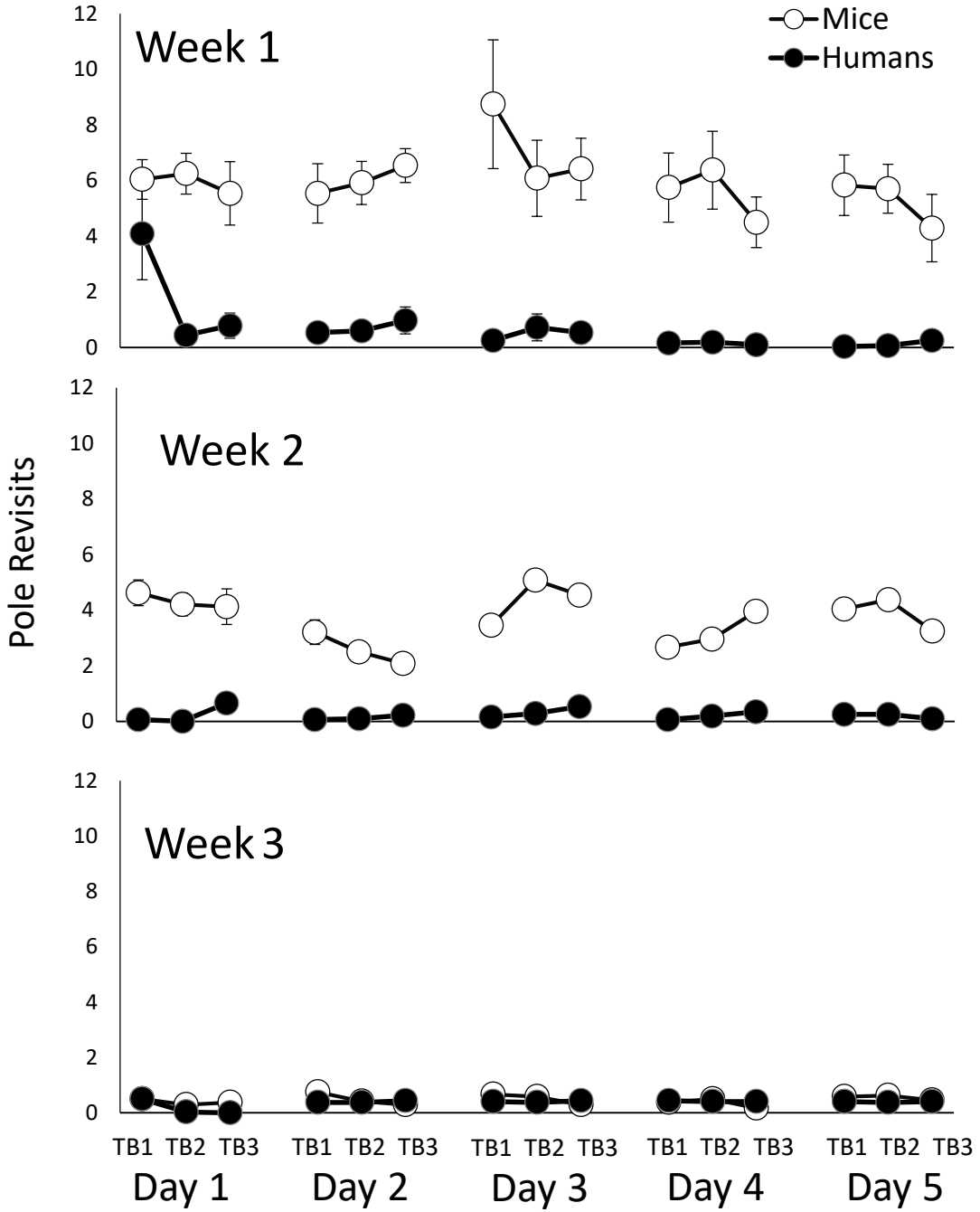

Week 1
Mice
Humans
Week 2
Week 3
Pole Revisits
12
10
8
6
4
2
0
TB1 TB2 TB3
Day 1
Day 2
Day 3
Day 4
Day 5

Fig 3

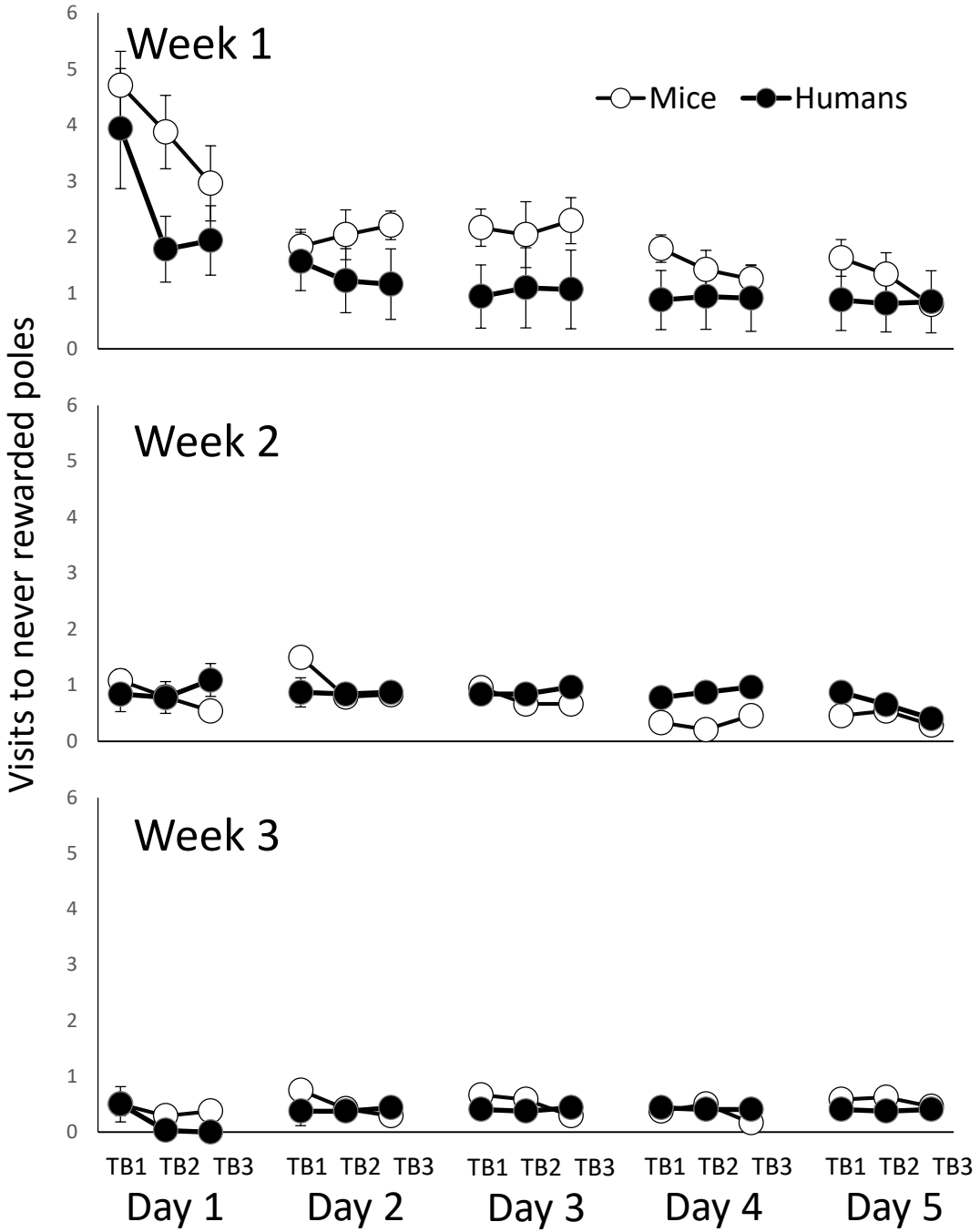
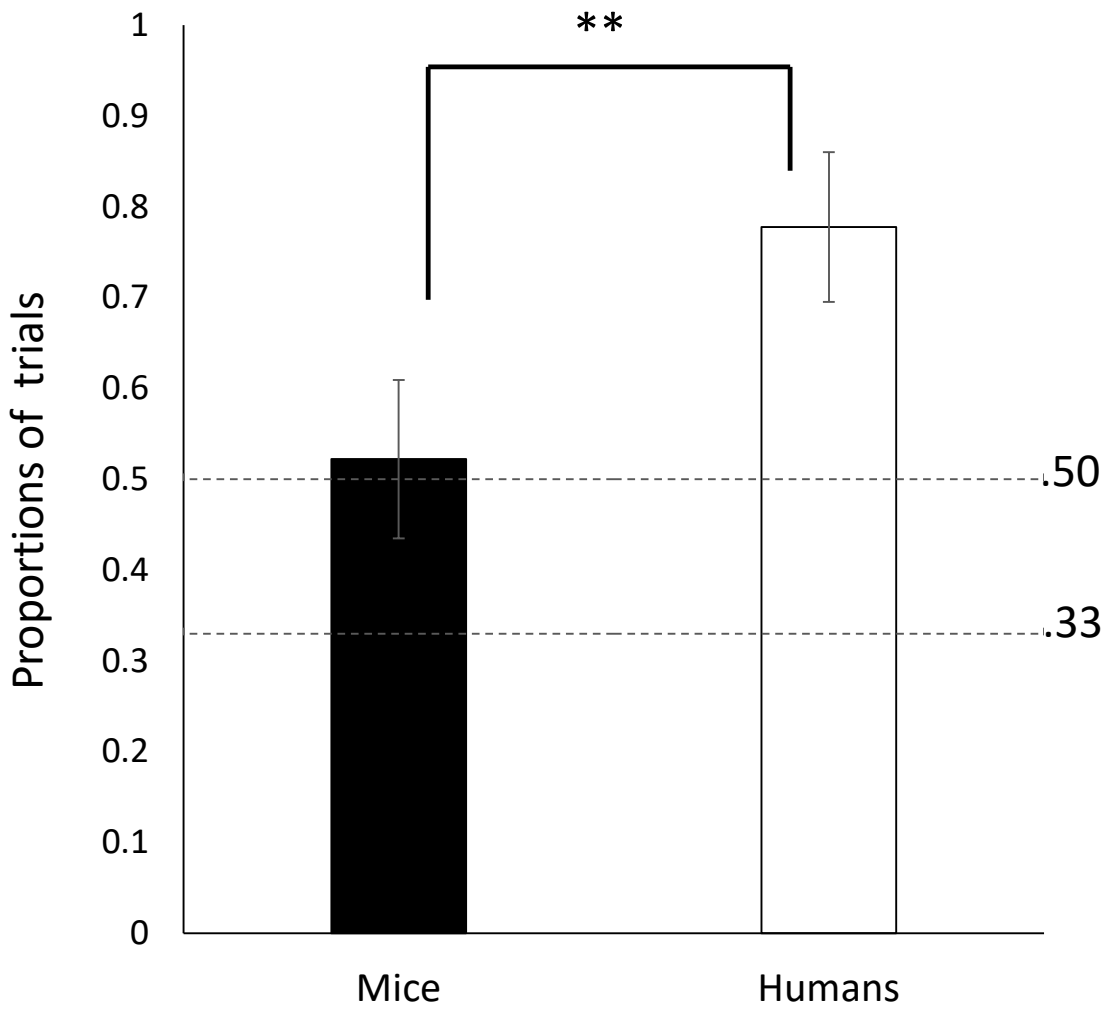

Week 1
Mice
Humans
Week 2
Week 3
Visits to never rewarded poles
0
1
2
3
4
5
6
TB1 TB2 TB3
Day 1
Day 2
Day 3
Day 4
Day 5

Fig 4

Fig 5


**
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportions of trials
.50
.33
Mice
Humans

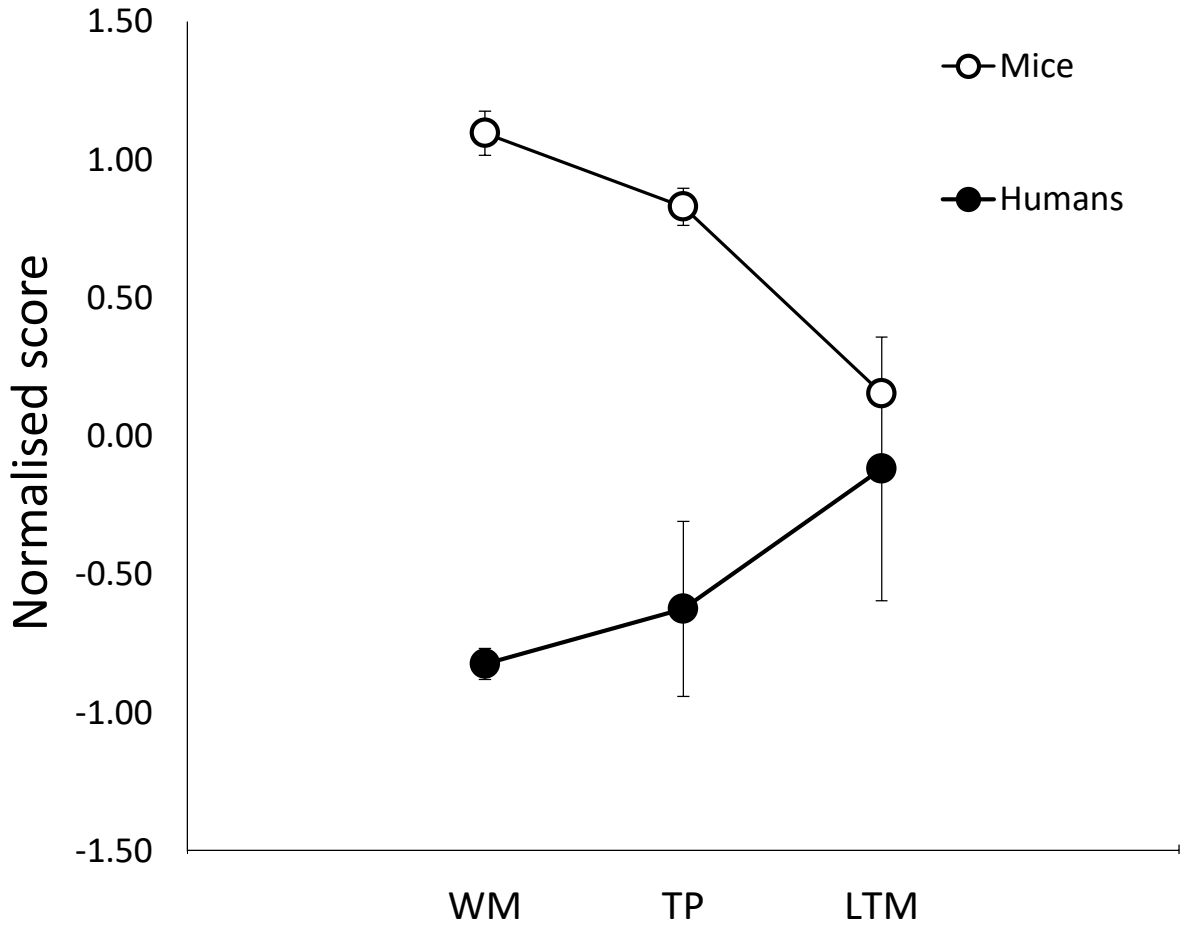

1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
Normalised score
Mice
Humans
WM
TP
LTM

Fig 6

Fig 7

Black Poles Rewarded

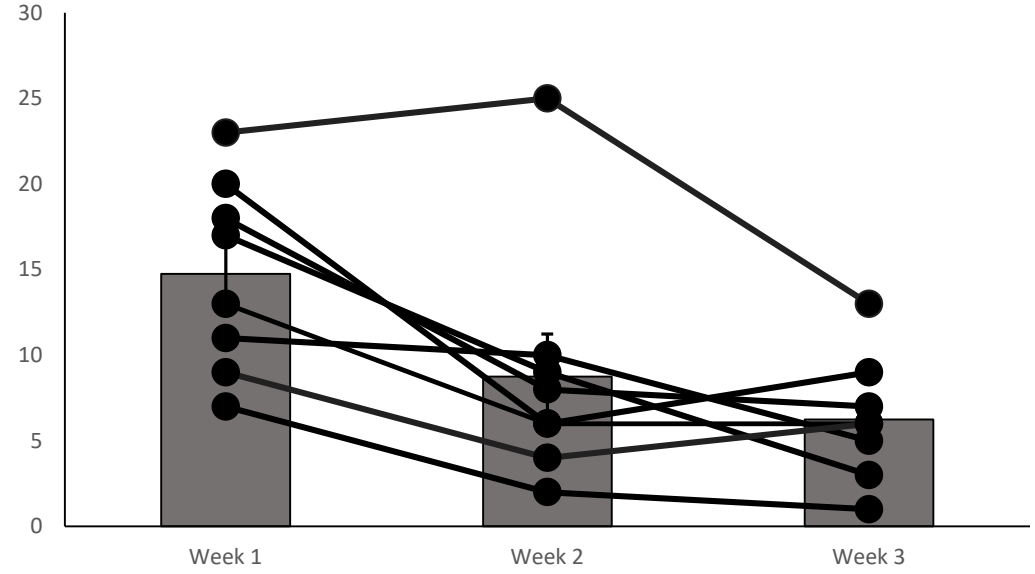
a)

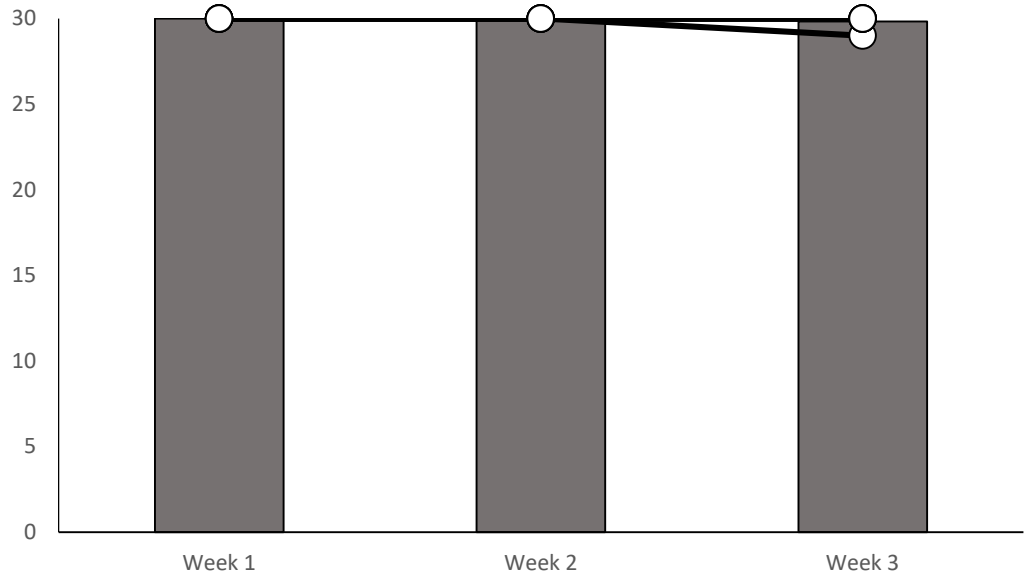
Humans

b)

Mice

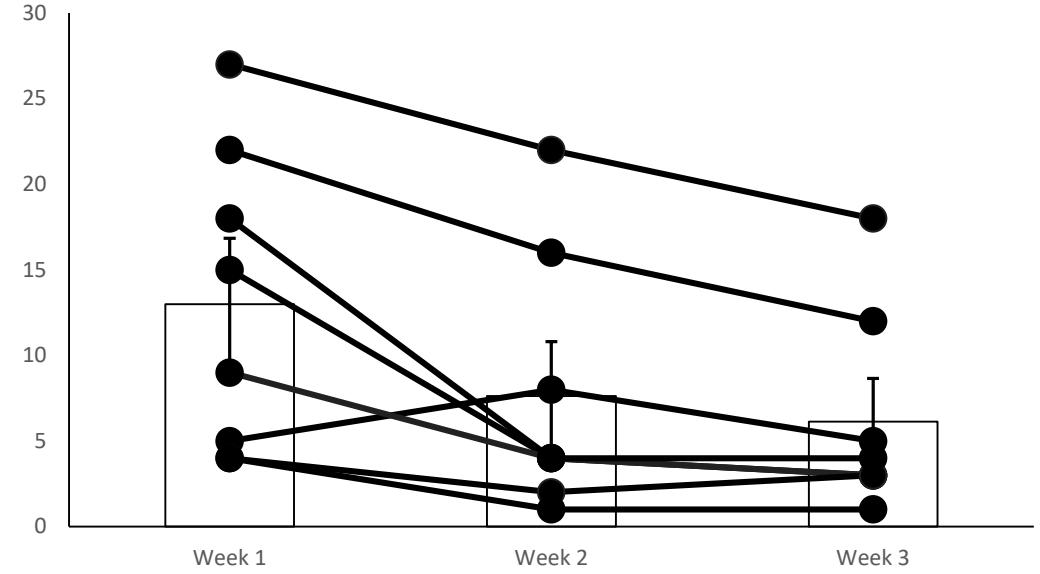
Number of different sequences





White Poles Rewarded

c)



d)