Behavioural Brain Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Research report

Negative transfer effects between reference memory and working memory training in the water maze in C57BL/6 mice

Lucas Ezequiel Serrano Sponton[a,b,*,1], Gonzalo Jose Soria[b,1], Sylvain Dubroqua[c,d], Philipp Singer[c,e], Joram Feldon[c], Pascual A. Gargiulo[b], Benjamin K. Yee[c,f,**]

[a] Department of Neurosurgery, Mainz University Hospital, Langenbeckstraße 1, 55131, Mainz, Germany
[b] Laboratorio de Neurociencias y Psicología Experimental, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Facultad de Ciencias Médicas, Universidad Nacional de Cuyo, 5500, Mendoza, Argentina
[c] Laboratory of Behavioural Neurobiology, ETH Zurich, Schorenstrasse 16, CH-8603, Schwerzenbach, Switzerland
[d] NYU-ECNU Institute of Brain and Cognitive Science, NYU Shanghai, Shanghai, China
[e] Roche Diagnostics, Hoffman-La Roche, CH-6343, Rotkreuz, Switzerland
[f] Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Mouse
Spatial learning
Transfer effect
Water maze

ABSTRACT

The water maze is one of the most widely employed spatial learning paradigms in the cognitive profiling of genetically modified mice. Oftentimes, tests of reference memory (RM) and working memory (WM) in the water maze are sequentially evaluated in the same animals. However, critical difference in the rules governing efficient escape from the water between WM and RM tests is expected to promote the adoption of incompatible mnemonic or navigational strategies. Hence, performance in a given test is likely poorer if it follows the other test instead of being conducted first. Yet, the presence of such negative transfer effects (or proactive interference) between WM and RM training in the water maze is often overlooked in the literature. To gauge whether this constitutes a serious concern, the present study determined empirically the magnitude, persistence, and directionality of the transfer effect in wild-type C57BL/6 mice. We contrasted the order of tests between two cohorts of mice. Performance between the two cohorts in the WM and RM tests were then separately compared. We showed that prior training of either test significantly reduced performance in the subsequent one. The statistical effect sizes in both directions were moderate to large. Although extended training could overcome the deficit, it could re-emerge later albeit in a more transient fashion. Whenever RM and WM water maze tests are conducted sequentially in the same animals – regardless of the test order, extra caution is necessary when interpreting the outcomes in the second test. Counterbalancing test orders between animals is recommended.

1. Introduction

The water maze is a common and robust test of hippocampus-dependent spatial learning in rodents [1]. Guided solely by distal extra-maze cues, the animals learn to navigate from any release point in the perimeter of a featureless circular pool of water to an escape platform hidden just underneath the water surface [1–3]. In his seminal paper, Morris [4] described in details the development and procedures to evaluate spatial reference memory (RM) with the location of the platform fixed to one location across trials and across days throughout. Learning is evident by efficient escape performance measured by the time or distance taken to reach the escape platform, directionality of the swim path, and the development of a search preference in the vicinity of the platform location. Morris [4] went on to describe a 'matching-to-sample' procedure in which the location of the platform varied from day to day. The platform position is thus unknown to the animals in the first trial on any given day. Across days, the animals showed clear evidence of saving from trial 1 to trial 2 even though the platform location differed from one day to the next. The 'matching-to-sample' (or 'matching-to-place') procedure has since been modified as a test of short-term, trial-dependent, working memory (WM) for rodents (e.g., [5]), to be contrasted with the processes underlying long-term, trial-independent memory evaluated in the RM test [2]. Both WM and RM versions of the water maze navigation task are widely used in the cognitive phenotyping of mutant mice, and it is not uncommon for them to be applied to the same animals (see Table 1). Researchers might

* Corresponding author at: Department of Neurosurgery, Mainz University Hospital, Langenbeckstraße 1, 55131, Mainz, Germany.
** Corresponding author at:Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong.
E-mail addresses: lucas.serrano@unimedizin-mainz.de (L.E. Serrano Sponton), benjamin.yee@polyu.edu.hk (B.K. Yee).
[1] Joint first authors.

https://doi.org/10.1016/j.bbr.2017.10.033
Received 26 July 2017; Received in revised form 27 October 2017; Accepted 31 October 2017

Table 1
A survey of published studies in mice that employed the water maze RM and WM tests consecutively in the same animals. We performed a PubMed search for mouse studies only (regardless of treatment or manipulation) that evaluated the RM and WM version of the water maze test in the same cohort of mice. Relevant articles were first identified by the following Boolean search terms: {("mouse" or "mice") and "working" and "reference" and "water maze"} on the 22nd June 2017. The articles were then screened to exclude studies in which (i) the two tests were performed in separate cohorts of mice, (ii) the subjects were rodent species other than mouse, (iii) "a radial arm water maze" was used to assess WM, or (iv) the protocols were controversial or inappropriate, e.g., "reversal learning" following RM acquisition training was treated as a WM test, or the use of a cued trial (instead of a hidden platform trial) in the first trial of a day in the WM test. The table shows that there are notable inconsistencies between studies in terms of days of testing, and the number of trials per day. The RM → WM sequence is clearly the prevalent choice in this selection of studies. Amongst these studies, the WM test took as little as three, but also as many as 21 days to complete. The time separating the two water maze tests also varies substantially. Not all had explicitly reported whether the two tests were performed in the same or two distinct testing rooms. We also noted whether there was evidence in support of a significant learning effect in the second test (mostly WM) present in the control subjects. Considerable differences exist among these studies, especially in the method of assessing WM performance. § indicates studies reporting a change of rooms between RM and WM tests. ‡ indicates studies reporting that the RM and WM test were conducted in the same room. All other studies did not explicitly specify room change. * refers to the same study by Lawson et al. [22] in which test orders were counterbalanced between subjects. " ! " refers to studies in which the statistical evidence for successful learning in the control group in the second test was considered weak or absent.

References/Reports on RM → WM	Background strain	Average group sizes	RM parameters: trials/days × days	WM parameters: trials/days × days	#days b/w tests	Control Group Performance in the second test (= WM)
[12] Malleret et al. 1999	129-Sv-ter	16.5	4 × 10	4 × 5	0	! Lack of significant improvement from trial 1 to 2, but significant across 1-4 trials
[27] Inman et al. 2000	C57BL/6	19	4 × 5	2 × 21	?	Significant improvement from trial 1 to 2
[13] Zeng et al. 2001	C57BL/6	15	4 × 10	(max 8) × 6	?	WM assessed by criterion of escape latency < 20s in 3 consecutive trials.
[22] Lawson et al. 2002 *	C57BL/6	11.5	3 × 4	3 × 4	1	Significant improvement from trial 1 to 2
[7] Buhot et al. 2003	C57BL/6	10.5	4 × 9	6 × 5	42	! Lack of significant improvement from trial 1 to 2
[14] Huang et al. 2003	C57BL/6	10.3	4 × 8	4 × 7	0	Significant improvement from trial 1 to 2
[15] Janus et al. 2004	C57BL/6	12	4 × 4	4 × 12	42	! Lack of significant improvement from trial 1 to 2, but significant across 1-4 trials
[8] Giménez-Llort et al. 2005	C57BL/6	15.7	4 × 5	4 × 8	7	! Lack of significant improvement from trial 1 to 2
[9] Liu et al. 2006	SAMR1/SAMP6	10	2 × 5	2 × 9	2	Significant improvement from trial 1 to 2
[10] Bour et al. 2008	C57BL/6	8	4 × 4	4 × 4	2	! Lack of significant improvement from trial 1 to 2 in female controls
[16] Zhou et al. 2009	C57BL/6	8	4 × 3	4 × 3	0	Significant improvement from trial 1 to 2
[17] Espallergues et al. 2010	129/Sv	8.5	(2∼3) × (5∼9)	5 × 3	?	! Lack of significant improvement from trial 1 to 2, but significant across 1-5 trials
[18] D'Agostino et al. 2012	C57BL/6	not stated	4 × 5	4 × 3	1	! Lack of significant improvement from trial 1 to 2, but significant across 1-4 trials
[19] Liu et al. 2013	ICR	13	4 × 5	5 × 3	1	Significant improvement from trial 1 to 2
[28] Singer et al. 2013 §	C57BL/6	23	2 × 5	2 × 8	7	Significant improvement from trial 1 to 2
[20] Xu et al. 2011	C57BL/6	7.5	4 × 8	4 × 8	?	No Trials effect because mice were placed on daily new platform followed by one search trial
[11] Torres et al. 2015	BALB/c	12	4 × (2∼5)	4 × (4∼5)	0	! Lack of significant improvement from trial 1 to 2, but significant across 1-4 trials
[21] Rahman et al. 2016	C57BL/6	12	(3∼4) × 5	(max 7) × 3	3	WM assessed by criterion of escape latency < 10s in 2 consecutive trials

References/Reports on WM → RM	Background strain	Average group sizes	WM parameters: trials/days × days	RM parameters: trials/days × days	#days b/w tests	Control Group Performance in the second test (= RM)
[22] Lawson et al. 2002 *	C57BL/6	11.5	3 × 4	3 × 4	1	! Lack of significant improvement across days
[23] Singer et al. 2009a ‡	C57BL/6	6	2 × 12	4 × 8	5	Significant improvement across days
[24] Singer et al. 2009b §	C57BL/6	9	2 × 27	4 × 8	7	Significant improvement across days

have opted for such a design with the intention to save on the number of animals, to allow within-subjects comparison between tests, or because the animals were difficult to breed or too costly to generate. However, it necessarily introduces a confound – namely, the order of tests, which has not received any serious attention in recent methodological reviews (e.g., [2,6]).

A brief survey of the relevant mutant mouse studies with a within-subject design reveals that the majority has elected to assess RM first, followed by WM (Table 1). This tacit convention is not based on any clear theoretical grounds. It also does not appear to minimize transfer effects. Closer examination of the control performance in this collection of studies suggests that RM training reliably retarded subsequent learning in the WM test. Successful acquisition of the WM task – as indexed by improvement from trial 1 to trial 2 – was either absent altogether [7–10] or evident only in swim distance but not escape latency [10,11]. The difficulty often led researchers to increase the number of trials (typically 4 per day, but could be up to 8 trials) to accumulate small increments between successive trials, even though this modification did not reliably lead to rapid learning from trial 1 to trial 2 [7,8,10–21]. Some even added a cued trial or simply placed the mice directly on the platform prior to the first hidden platform trial [13,20].

By contrast, very few studies had employed a WM → RM within-subject design in mice. Lawson et al. [22] had carefully balanced the test order and reported that prior WM training did not retard learning in the succeeding RM task. This conclusion is undermined, however, by the fact that mice in the WM → RM order did not perform significantly above chance in the WM test (i.e., trials 2–3 preformance remained similar to trial 1), which incidentally lasted only for four days (see their Fig. 4). We also did not observe any difficulty in switching from WM to RM test in control mice [23,24]. When both WM and RM tests took place in the same room, only a tentative trend of a negative transfer was visible on the second day into the RM test: control performance was poorer than on the first day [23] when it was still procedurally indistinguishable from previous WM training. The change required by the RM test – namely, the constancy of platform location across days – only becomes apparent from the second day onwards. This transient effect was no longer visible when the RM test was conducted in another room, despite extensive prior WM testing for 27 days [24]. However, we did not control for test order in our experiments, and the control mice had performed poorly in the WM test. Hence, there is insufficient data to decide whether a WM → RM design may carry a less deleterious transfer effect or not.

One may suspect that RM training (with the never-changing platform location) could undermine the use of flexible memory demanded by the daily switching of platform location during the WM test. The poorer WM performance in RM-trained animals is attributed to greater sensitivity to memories of platform locations in previous days – which no doubt contributes to successful RM performance. Conversely, it is conceivable that WM-training may similarly impair subsequent RM performance. Under the WM procedure, the platform locations on previous days are irrelevant to the solution of the current day, and this might bias the animals against forming any long-term persistence for one specific location accumulated across days. This is in spite of evidence that rats trained with the 4-trial per day WM protocol exhibited a search preference near previous day's platform location [25,26] even when the previous platform location was learned 10 days ago [25]. The fact that such a win-stay tendency across consecutive test days could be demonstrated implies that the trial 1 performance (when the platform location is not known) was largely guided by the memory of the most recent platform location known to the animals rather than by more remote memories of platform locations experienced before. The "win-stay" tendency developed under WM training thus bears the hallmark of a recency effect, which biases the recall of more recent events – an essence of any delay matching task. This is in keeping with our hypothesis that WM training could potentially undermine the building up of long term persistence that supports strong RM performance, especially in the probe test that follows. It is therefore conceivable that prior WM training could disadvantage the animals when switching afterwards to the RM test in comparison with animals that had never before experienced WM training. It follows that the transfer effect attributable to the incompatible rules between the RM and WM procedures should be symmetrical: from RM to WM, and vice versa. However, this account cannot readily predict whether the effect would be stronger or more persistent in a particular test order. The present study was designed to empirically determine the magnitude of the transfer effect between RM and WM tests in the water maze in wild type C57BL/6 mice – the most common background strain of engineered mice. Two cohorts of adult males were evaluated across the RM and WM tests in succession, but with opposite orders. The RM and WM tests were conducted in two distinct rooms so as to ensure that any transfer or interference effect observed is not stemming from the use of a common set of spatial cues across tests.

2. Methods & materials

2.1. Subjects

Twenty-five 12-week old male mice were obtained from our in-house breeding facility (Laboratory of Behavioural Neurobiology, Swiss Federal Institute of Technology, Zurich). They were collected from multiple breeding pairs (C57BL/6NCrl) obtained from Charles River Laboratories (Sulzfeld, Germany). The mice were housed 2–4 to a cage in a temperature (22 ± 1 °C) and humidity-controlled (55 ± 5%) vivarium under a reversed light-dark cycle (lights on 1900–0700 h) and allowed free access to food and water throughout the study. All tests were conducted during the dark phase. All experiments described here had been approved by the Cantonal Veterinary Office in Zurich in accordance with the Swiss Animal Welfare Act and Ordinance. Animal husbandry and experimentation were in keeping with the standards stipulated by the European Commission Directive 2010/63/EU and the "Principles of Laboratory Animal Care" (NIH publication No. 86-23, revised 1985). All efforts had been made to minimize any potential suffering of the animals and the number of animals required. The animals were randomly allocated to one of two groups according to whether they were trained first with the working memory test followed by reference memory test (WMRM, N = 12), or vice versa (RMWM, N = 13).

2.2. Apparatus

Two water mazes identical in size and construction were adopted here from our previous studies [23,24,29]. They were made of fibreglass and painted white, 101 cm in diameter and 36 cm high. Fresh tap water (25 ± 1 °C) filled the water maze to a depth of 19 cm. A solid transparent Plexiglas cylinder 9.5 cm in diameter and 18.5 cm high served as the escape platform in the WM and RM training. A smaller 7-cm diameter platform was used for pre-training and the cued task (see Procedures). The top of the escape platform was submerged 0.5 cm below the water and therefore not visible to the animals. The two water mazes were positioned in the middle of two distinct rooms (Room 1 and Room 2) with 20–25 lx ambient illumination; and each room was enriched with a unique set of extramaze cues. The swim paths were tracked by a digital camera connected to a PC running Ethovision 3.1 (Noldus Technology, the Netherlands), which also extracted all dependent measures: escape latency, distance traversed (or path length), and time or distance spent in defined zones (as percent of total) in all probe tests.

2.3. Procedures

2.3.1. Pre-training

This was performed in a separate room. All animals were first familiarized with being put in water and the opportunity to escape from the water by climbing onto a nearby stable platform [30]. Each animal was gently released into a bucket (34 cm in diameter, and 37 cm tall) filled with water (25 ± 1 °C) to a depth of 19 cm with its head pointing towards an escape platform. The platform was 7 cm in diameter, 18.5 cm tall, and had a carpeted top surface (3M™ Nomad™ Scraper Matting). The platform was always placed in the centre of the bucket, and the animals were allowed 60 s to climb onto the escape platform. They were allowed 10 s on the platform before being removed and immediately released again into the water as before. All mice readily climbed up, and stayed on, the escape platform. None of them showed any motor disturbances. The cued task began on the next day in the testing room where they would eventually receive their first spatial learning test (see Fig. 1).

2.3.2. Cued task

The cued task evaluated performance of the water escape task in the absence of an explicit demand on allocentric spatial memory [23]. The 7-cm diameter escape platform was used, and a salient local cue in the form of a three-dimensional arrow (15 cm long, 4 cm wide) was mounted 12 cm directly above it. The platform was placed randomly amongst 8 possible positions: 10.5 cm from the maze wall in the directions of N, E, S, W, NE, SE, SW and NW. A trial began by releasing the mouse in the centre of the water maze. A maximum of 60 s was allowed to locate the platform, otherwise the animal was guided to the platform by the experimenter, and an escape latency of 60 s assigned. The mouse was allowed 20 s on the platform before being retrieved and dried. It was then kept in a waiting cage for 120 s before the second trial began with the cued platform moved to another random location. The WMRM and RMWM groups achieved comparable performance after 2 and 4 days, respectively, in this non-spatial navigational task (see Section 3.1).

2.3.3. Working memory task

The animals were required to learn the novel position of the platform revealed to them in trial 1 on each day in order to efficiently navigate to the same location (i.e. matching) in trial 2 on the same day after some delay (or inter-trial interval). The daily change of platform location ensured that the solution on a given day was irrelevant for the next day, thus taxing flexible working memory (short-term retention and trial-dependent retrieval). The procedures of WM training were similar to the cued task except that the platform was 9.5 cm in

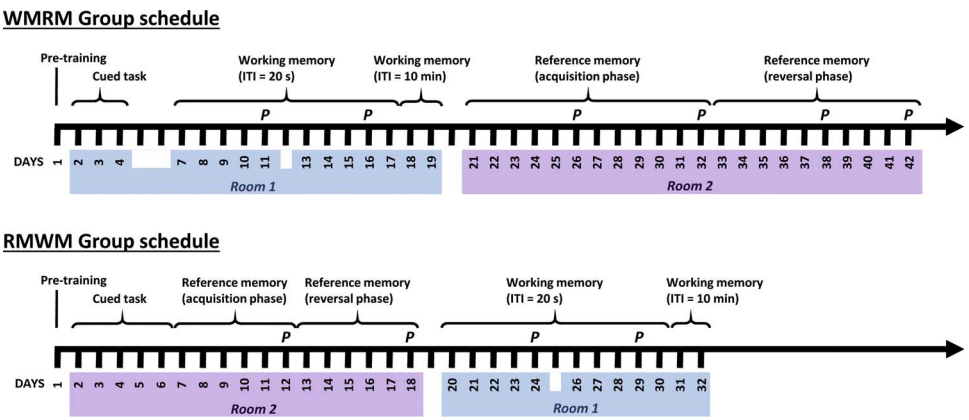

Fig. 1. The training schedules between the two groups of mice differed primarily in terms of whether working memory training was performed before reference memory training (WMRM group) or vice versa (RMWM group). Within each day, two consecutive trials were performed, separated by a pre-established inter-trial interval (ITI). "*P*" indicates days on which a probe test was conducted (see *Procedures*). Comparisons between the two groups on the same test (performed in the same room) permitted us to examine the transfer effects between reference memory training and working memory training.

diameter, submerged without any local cues, and its position remained unchanged from trial 1 to trial 2 on the same day but varied between days amongst 12 pre-determined locations that followed the twelve clock positions: 10.5 cm off the wall at 2, 4, 6, 8, 10 and 12 o'clock, and 26 cm off the wall at 1, 3, 5, 7, 9 and 11 o'clock [23]. Each position was used once according to a pseudorandom sequence across 12 test sessions (see Fig. 1). Similarly, the start positions were also randomized amongst the twelve clock positions with the restriction that they never repeated across the two trials on the same day. This applies to days on which the platform was in place in both trials (days 7–10, 13–15, 17–18 in Fig. 1). On the two probe test days, the first trial was as described with randomized start position, but the start position of the second trial (the probe test itself) was always at 180° from the target position (see later). The sequence of platform locations across days as well as the start positions on all trials were uniquely and independently determined for each animal. The animals were allowed a maximum of 60 s to locate the escape platform and 20 s on the platform. The inter-trial interval (ITI) between trial 1 and trial 2 was 20 s plus a few seconds for the experimenter to quickly dry the mouse and move to the next start position. For practical purpose, we referred to this minimal ITI as "20s". In the last two days, the ITI was extended to 10 min, which referred to the duration the mouse was kept in a waiting cage after it had spent 20 s on the platform. The two groups (WMRM and RMWM) underwent the same schedule of WM test over a period of 13 days as depicted in Fig. 1. A probe test was performed after 4 training sessions followed by a day-off, and then another probe test was performed after 3 more training sessions. In the probe test, trial 1 was as described above, but trial 2 was conducted without the platform – which constituted a probe test, in which the animals were allowed to search in the water maze for 40 s. Spatial bias during the probe trial was evaluated by dividing the circular maze area into 6 equal sectors, with the target sector centring on where the platform was positioned in trial 1. The time and distance spent in the target sector were contrasted with the time and distance recorded in the other five sectors (see Section 3.2.2).

2.3.4. Reference memory task

The RM test comprised two phases: (*i*) acquisition, with escape platform fixed in one location throughout, and (*ii*) reversal learning, with the escape platform moved 180° to the middle of the opposite quadrant. Within each phase, the platform location was fixed across days and between trials. Each animal was randomly assigned a platform location (middle of the NE, SE, SW or NW quadrants) prior to the acquisition phase, and a non-repeating random sequence of start positions (N, E, S or W) each day. On each trial, the mice were allowed 60 s to locate the platform, or else they would be guided to the platform by the experimenter. After spending 20 s on the platform, the mouse was picked up and dried before the next trial. The ITI was thus 20 s plus a few seconds needed for the experimenter to quickly dry the mouse and move to the next start position (i.e., in a manner that was identical to the 20-s ITI condition in the WM training procedure described above). Here, there were only 2 trials per day so the amount of training per day was also kept identical to that in the WM protocol. A probe test was performed on selected days (as indicated in Fig. 1), in which the platform was removed, and the animal was allowed to search for 40 s after being released from the point opposite to the target (i.e., again in a manner that was identical to the probe tests performed in the WM training procedure). Spatial bias towards any particular quadrants and the persistence of the bias were quantified in a probe test. The target quadrant was always defined as the quarter-circle centred on the platform location. The remaining three quadrants (two adjacent and one opposite) of equal area were defined relative to the target quadrant. The training schedule differed somewhat between the two groups (Fig. 1). The RMWM group first underwent 5 days of acquisition and a probe test the following day. This was followed by 5 days of reversal learning and another probe test. The WMRM group underwent an extended acquisition phase lasting for 12 days, with a probe test conducted on the 6th and 12th day. This was followed by reversal learning that lasted 10 days in total, with a probe test conducted on the 6th and 10th day into the reversal phase.

3. Results

3.1. The cued task

The initial performance of the RMWM group was somewhat unstable compared with the WMRM group. We therefore extended the cued task training to four days in the RMWM group. By the last two days, the two groups had achieved comparable performance (Fig. 2). Separate 2 × 4 (group × trials) split-plot ANOVAs of escape latency and path length did not reveal any statistically significant effect. Swim speed was also stable without any suggestion of a difference between groups.

3.1.1. Prior reference memory training impaired subsequent working memory performance

Working memory performance was indexed by the improved efficiency to locate the escape platform from trial 1 (when the location of the platform was not known) to trial 2. This was clearly evident across the 8 days of testing with 20-s ITI (i.e., minimal delay), but it was markedly weaker in the RMWM group in comparison with the WMRM group (Fig. 3A–D). A 2 × 8 × 2 (group × days × trials) ANOVA of escape latency yielded a significant group × trials interaction [$F_{(1,23)} = 5.85$, $p < 0.05$, $\eta_p^2 = 0.20$] that accompanied the main effect of trials [$F_{(1,23)} = 31.21$, $p = 0.00001$, $\eta_p^2 = 0.58$] (Fig. 3A and C). A similar pattern of outcomes was obtained in the path length analysis (Fig. 3B and D), although the critical group × trials interaction failed to achieve statistical significance [$F_{(1,23)} = 2.98$, $p = 0.098$, $\eta_p^2 = 0.16$; trials effect: $F_{(1,23)} = 35.90$, $p < 0.00001$, $\eta_p^2 = 0.61$]. To better

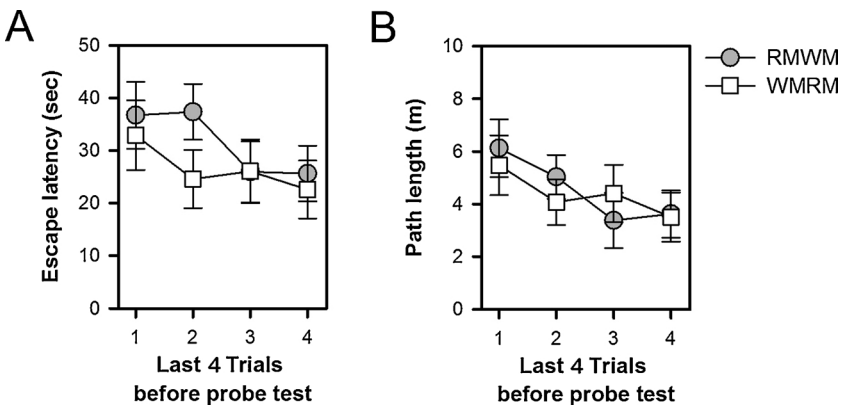

Fig. 2. Performance in the cued task. Escape latency (A) and path length (B) to reach the platform over the last 4 trials were comparable between groups. All values shown express mean ± SEM.

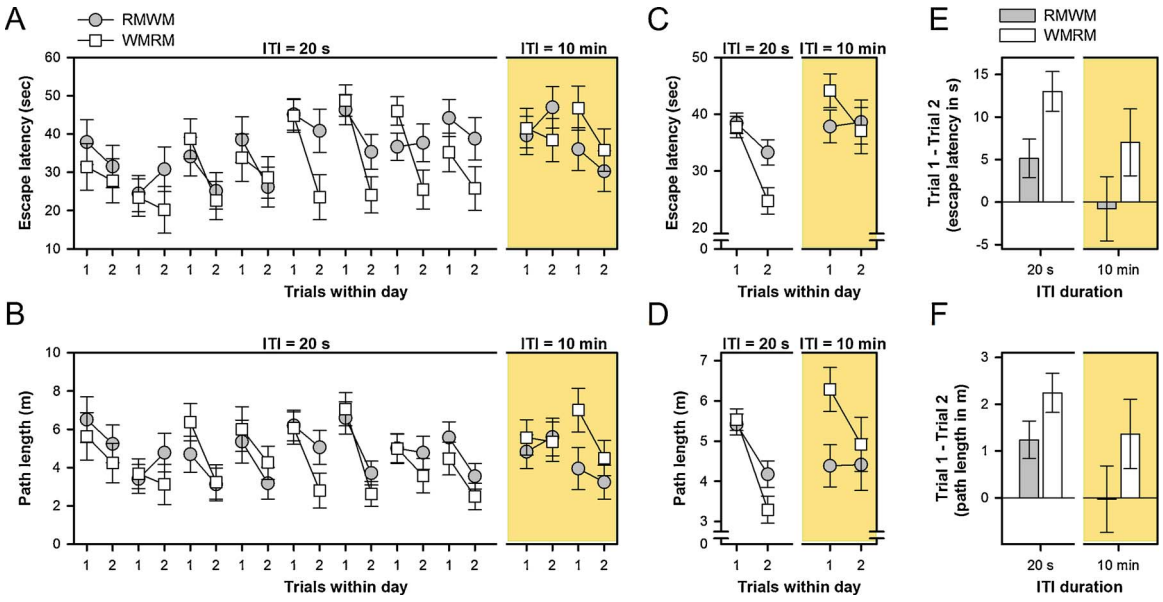

Fig. 3. Working memory performance in the water maze was indexed by escape latency and path length, expressed as a function of trials, across 10 days (A and B, respectively). Performance across the 8 days in which the ITI from trial 1 to trial 2 was 20 s was collapsed across days and depicted in the *left* panel of (C) and (D), respectively. Performance across the last 2 days in which the ITI from trial 1 to trial 2 was 10 min was collapsed across days and depicted in the *right* panel of (C) and (D), respectively. To better visualize the critical interaction involving the factor trials shown in (C) and (D), we also plotted the difference scores contrasting explicitly the saving from trial 1 to trial 2, calculated daily for both escape latency (E) and path length (F) for each animal. These two plots effectively transform the *trials* effect in the relevant ANOVAs into the dependent measure that indexes *trial-dependent saving*. The depiction of the *group* effect in (E) and (F) corresponds to the depiction of the *group* × *trials* interaction in (C) and (D), respectively. The *group* × *ITIs* interaction depicted in (E) and (F) is in effect with the *group* × *ITIs* × *trials* interaction depicted in (C) and (D), respectively. The corresponding *p*-values are therefore identical. All values shown express mean ± SEM.

visualize the critical interaction involving the factor trials, we also plotted the difference scores contrasting explicitly the saving from trial 1 to trial 2 (calculated daily for both escape latency and path length for each animal) as depicted in Fig. 3E and F.

When the ITI was extended to 10 min in the last two test days, an improvement from trial 1 to trial 2 was again visible in the WMRM group but hardly discernible in the RMWM group (see Fig. 3A–D). However, the critical group by trials interaction was not statistically significant [escape latency: $F_{(1,23)} = 2.05$, $p = 0.17$, $\eta_p^2 = 0.08$; path-length: $F_{(1,23)} = 1.69$, $p = 0.21$, $\eta_p^2 = 0.07$] in separate 2 × 2 × 2 (group × days × trials) ANOVAs of escape latency and path length.

An additional 2 × 2 × 2 (group × ITIs × trials) ANOVA that included the escape latency data from both ITI conditions again yielded a significant group × trials interaction [$F_{(1,23)} = 4.31$, $p < 0.05$, $\eta_p^2 = 0.16$] and a main effect of trials [$F_{(1,23)} = 5.84$, $p < 0.01$, $\eta_p^2 = 0.27$] but no evidence for a 3-way interaction [$F = 0.00$]. The negative transfer effect of prior RM training thus appeared to be independent on the delay demand of the WM test. The path length analysis revealed a similar pattern of results albeit lacking somewhat in statistical power: the critical group × trials interaction was only close to statistical significance [$F_{(1,23)} = 3.28$, $p = 0.08$, $\eta_p^2 = 0.13$] despite the presence of a trials effect [$F_{(1,23)} = 10.45$, $p < 0.005$, $\eta_p^2 = 0.31$] (also see Fig. 3E and F).

The two groups did not differ in swim speed throughout the WM test.

3.1.2. The probe tests during WM testing

Two probe tests were conducted over the course of WM testing (see Fig. 1) when the platform was removed from the water maze in trial 2 and we recorded the search pattern of each animal over 40 s. We divided the maze surface into six equal sectors with the target sector centring on the platform location presented in trial 1 just prior to the probe test. Our attempts to evaluate (trial 2) performance by means of a probe test were not successful and did not identify any group difference. The animals did not show any overall preference for the target sector and neighbouring areas (Fig. 4A–B). Over the first 20 s, the animals were spending more time and distance in areas away from the target sector (Fig. 4C–D). This pattern likely stemmed from our choice of releasing the animals from the furthest point away from the target. No spatial bias was evident by the last 20 s. Separate 2 × 2 × 4 × 6

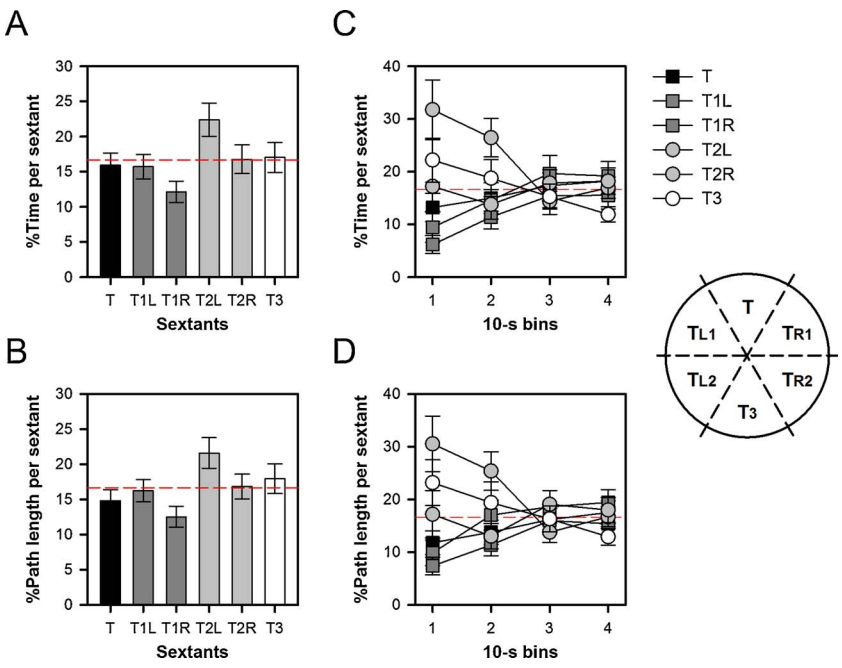

Fig. 4. Two probe tests were conducted during the working memory test. The procedure was identical to any other working memory day, except that the platform was removed during trial 2, and the animals were allowed to explore for 40 s. The search pattern was analysed by dividing the maze surface into six sectors, with the target sector (T) centring on where the platform was positioned in the preceding trial 1. The position of the two sectors to the left (TL1 and TL2), and the two to the right (TR1 and TR2), of the target sector (T), as well as the sector opposite to T, are depicted. The search pattern was examined by %time and %path length in (A) and (B), respectively, recorded over 40s. In addition, the evolution of the two measures across the four successive 10-s bins (spanning across the total test duration of 40 s) was depicted in (C) and (D). All values shown express mean ± SEM. Dashed horizontal lines indicate chance level performance.

(group × tests × 10-s bins × sectors) ANOVAs of the percent time and percent path length only revealed a significant effect of sectors [%time: $F_{(5,115)} = 2.41$, $p < 0.05$, $\eta_p^2 = 0.10$; %path-length: $F_{(5,115)} = 2.39$, $p < 0.05$, $\eta_p^2 = 0.09$] and the sectors × bins interaction [%time: $F_{(15,345)} = 3.30$, $p < 0.00005$, $\eta_p^2 = 0.13$; %path-length: $F_{(15,345)} = 3.19$, $p < 0.0001$, $\eta_p^2 = 0.12$].

Our animals did not show any overall bias for the target sector on probe tests during WM testing (Fig. 4). We do not have any ready explanation for the observed preference for TL2 and relative avoidance of TR1. This was unexpected although no comparable studies in mice have attempted such probe tests. Nonetheless, a clear spatial bias for areas near the target has been reported in rats undergoing WM training in a probe test conducted in the first trial [25,26,31,32]. Although these reports evaluated the preference for yesterday's platform location, their findings in rats suggest that WM training (at least in a 4-trial-a-day protocol) can instil a win-stay strategy generalizable across days – and thus making our failure to show a preference for today's platform location on trial 2 more astounding. Further studies are clearly warranted. Nonetheless, the absence of the expected target preference in both RMWM and WMRM groups is noteworthy here, because this was clearly independent of whether the mice had been previously exposed to the RM protocol or not.

3.2. Prior working memory training retarded subsequent reference memory learning

3.2.1. Acquisition phase

Over the first five days of acquisition training, the WMRM group generally performed less well than the RMWM group (Fig. 5A–B). A 2 × 5 × 2 (group × days × trials) ANOVA yielded a significant group effect in terms of path length [$F_{(1,23)} = 4.37$, $p < 0.05$, $\eta_p^2 = 0.16$] but not of escape latency [$F_{(1,23)} = 1.34$, $p = 0.26$, $\eta_p^2 = 0.06$]. The latency measure did not reveal a group difference primarily because the WMRM group was swimming significantly faster ($0.183 \pm 0.008\ ms^{-1}$) compared with the RMWM group ($0.156 \pm 0.008\ ms^{-1}$) [$F_{(1,23)} = 5.57$, $p < 0.05$, $\eta_p^2 = 0.20$]. The speed difference was most pronounced on the first day (RMWM = $0.154 \pm 0.009\ ms^{-1}$ vs WMRM = $0.205 \pm 0.009\ ms^{-1}$), and largely disappeared by the fifth day (RMWM: $0.170 \pm 0.011\ ms^{-1}$ vs WMRM: $0.176 \pm 0.012\ ms^{-1}$). The day dependency of the speed difference was confirmed by the significant group × days interaction [$F_{(4,92)} = 2.49$, $p < 0.05$, $\eta_p^2 = 0.10$].

The probe test conducted on the following day confirmed the impression that prior WM training had influenced current RM performance (Fig. 5C–D). A 2 × 4 × 4 (group × 10-s bins × quadrants) ANOVA of %time per quadrant yielded a significant group × quadrants interaction [$F_{(3,69)} = 3.24$, $p < 0.05$, $\eta_p^2 = 0.12$]. The analysis of% path length revealed a comparable impression even though the interaction term only approached statistical significance [$F_{(3,69)} = 2.48$, $p = 0.07$, $\eta_p^2 = 0.10$]. One-sample *t*-tests confirmed that the preference for the target quadrant (%time and %path length) differed significantly from chance level (25%) in the RMWM group ($p < 0.05$), but not in the WMRM group (Fig. 5C–D). Supplementary analyses restricted to the target quadrant showed that the group difference was relatively stable across the four 10-s bins (Fig. 5C'–D'), yielding a significant group difference in both measures [%time in target quadrant: $F_{(1,23)} = 7.15$, $p < 0.05$, $\eta_p^2 = 0.24$; %path length in target quadrant: $F_{(1,23)} = 5.03$, $p < 0.05$, $\eta_p^2 = 0.18$], without evidence for an interaction with time [$F < 1.0$, *ns*].

In an attempt to match performance in the WMRM group to that of the RMWM group before the start of reversal learning, we provided 5 extra days of training (followed then by another probe test). The WMRM group showed marked improvement (Fig. 5A–B). Statistical comparison (5-day blocks × days × trials ANOVA) confirmed the presence of an improvement from the first to the last block in the WMRM group [latency: $F_{(1,11)} = 15.30$, $p < 0.005$, $\eta_p^2 = 0.58$; path-length: $F_{(1,11)} = 17.12$, $p < 0.005$, $\eta_p^2 = 0.60$]. In the probe test that followed, a clear preference for the target quadrant emerged in the WMRM group (Fig. 5E–F), approaching the pattern exhibited by the RMWM group in the first probe test before. Separate 4 × 4 (quadrants × 10 s bins) repeated measures ANOVAs of %time and %path length yielded a significant effect of quadrants [%time: $F_{(3,33)} = 5.26$, $p < 0.005$, $\eta_p^2 = 0.32$; %path length: $F_{(3,33)} = 6.71$, $p = 0.005$, $\eta_p^2 = 0.38$] and its interaction with bins [%time: $F_{(9,99)} = 2.98$, $p < 0.005$, $\eta_p^2 = 0.21$; %path length: $F_{(9,99)} = 3.36$, $p = 0.005$, $\eta_p^2 = 0.23$]. The interaction was due to the gradual reduction of target quadrant preference over the four bins (data not shown). One-sample *t*-tests further confirmed that the WMRM mice now exhibited an above chance preference for the target quadrant ($p < 0.05$) in both %time and %path length measures (as indicated in Fig. 5E–F). Direct comparison between probe 1 performance by the RMWM group and probe 2 performance by the WMRM group only yielded a highly significant quadrants effect [%time: $F_{(3,69)} = 12.08$, $p < 0.001$, $\eta_p^2 = 0.34$; %

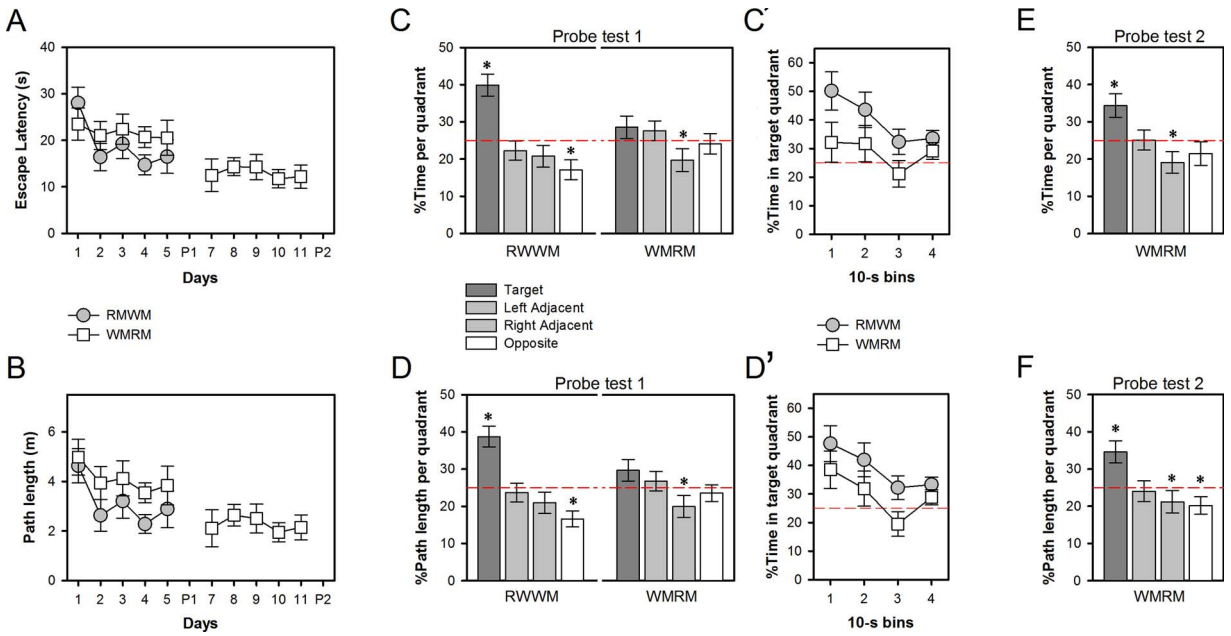

Fig. 5. Performance during the acquisition phase of the reference memory test is illustrated by the reduction of escape latency and path length over successive days as shown in (A) and (B), respectively. After 5 days, a probe test was performed in which the escape platform was removed and the search pattern indexed by % time and %path-length over the 40 s of the test was analysed (C and D). The evolution of the time or path length spent inside the target quadrant across successive 10-s bins was subjected to further analysis (C' and D'). The WMRM group was given five additional days of "remedial" acquisition training as before, followed by another probe test 24 h later. The outcomes of this probe test (by the WMRM group only) are shown in E and F. The extra training was effective in improving probe test performance in the WMRM group to a level comparable to that of the RMWM earlier. All values are expressed as mean ± SEM. Horizontal dashed lines indicate chance level performance. * in (C, C', D, D', E and F) indicates significant deviation from chance level performance based on separate one-sample *t*-tests ($p < 0.05$, two-tailed).

path length: $F_{(3,69)} = 13.14$, $p < 0.001$, $\eta_p^2 = 0.36$], but not any significant group effect or its interactions [all *F*'s < 1.0,*ns*].

3.2.2. Reversal phase

Switching the escape platform to the opposite quadrant 48 h after the last acquisition training led to comparable disruption of performance in both groups (Fig. 6A–B). However, as reversal learning progressed, the RMWM group gradually outperformed the WMRM group. This impression was confirmed by the presence of a significant group × days interaction [$F_{(4,92)} = 2.83$, $p < 0.05$, $\eta_p^2 = 0.11$] in a 2 × 5 × 2 (group × days × trials) ANOVA of escape latency. Although the interaction term did not reach statistical significance in terms of path-length [$F_{(4,92)} = 2.28$, $p = 0.07$, $\eta_p^2 = 0.09$], the pattern was highly similar (Fig. 6A–B). It thus appeared that prior working memory training (over 2 weeks ago) had continued to influence RM learning at this stage despite effective extra acquisition training. This impression was, however, only tentatively supported by the probe test that followed.

An overall preference for the target quadrant was visible in both the RMWM and the WMRM groups (Fig. 6C–D), although this preference was visually more pronounced in the RMWM group. The 2 × 4 × 4 (group × 10-s bins × quadrants) ANOVAs clarified that the difference in search pattern was statistically tentative. The group × quadrants interaction only approached significance in [%time: $F_{(3,69)} = 2.35$, $p = 0.08$, $\eta_p^2 = 0.09$; % path length: $F_{(3,69)} = 2.14$, $p = 0.10$, $\eta_p^2 = 0.09$]. Additional analyses restricted to the new target quadrant across four 10-s bins also did not reveal any significant group difference (Fig. 6C'–D'). Supplementary one-sample *t*-tests further confirmed that the time spent in the target quadrant was significantly above chance ($p < 0.05$) in both groups in terms of %time and %path length (as indicated in Fig. 6C and D).

Next, the WMRM group resumed reversal learning for another three days before being subjected to another probe test in an effort to test whether further improvement was feasible. Performance rapidly and substantially improved in these three days of remedial training, at a level matching the final performance (escape latency or path length) by the RMWM group (Fig. 6A–B). The probe test that followed strengthened this impression (Fig. 6E–F). Separate 4 × 4 (quadrants × 10 s bins) repeated measures ANOVA of %time and %path length yielded a significant effect of quadrants [%time: $F_{(3,33)} = 3.86$, $p < 0.05$, $\eta_p^2 = 0.26$; %path length: $F_{(3,33)} = 5.63$, $p < 0.005$, $\eta_p^2 = 0.34$] but not the quadrants by bins interaction [$F < 1$, *ns*]. One-sample *t*-test confirmed that preference for the target quadrant remained significantly above chance ($p < 0.05$) in both measures (as indicated in Fig. 6E–F).

Again, we directly contrasted the performance of the WMRM group in the second probe test (post-reversal learning) against the performance of the RMWM group in the first probe test (post-reversal learning) in a 2 × 4 × 4 (group × 10-s bins × quadrants) ANOVA. The analysis only yielded a significant quadrants effect [%time: $F_{(3,69)} = 7.06$, $p < 0.005$, $\eta_p^2 = 0.24$; %path length: $F_{(3,69)} = 9.37$, $p < 0.001$, $\eta_p^2 = 0.29$]. Neither the group by quadrants interaction nor the three way interaction achieved statistical significance [all *F*'s < 1,*ns*]. Hence, the additional reversal training given to the WMRM animals effectively compensated the initial performance deficit in the first post-reversal learning probe test.

Finally, we compared performance in the two (post-reversal learning) probe tests in order to examine if performance has significantly improved in the WMRM group. These ANOVAs only yielded a highly significant effect of quadrants [%time: $F_{(3,33)} = 5.49$, $p < 0.005$, $\eta_p^2 = 0.33$; %path length: $F_{(3,33)} = 8.11$, $p < 0.001$, $\eta_p^2 = 0.42$]. There was no suggestion for an improvement due to the lack of a quadrants by tests interaction [%time: $F_{(3,33)} = 1.63$, $p = 0.20$, $\eta_p^2 = 0.13$; %path length: $F_{(3,33)} = 1.66$, $p = 0.19$, $\eta_p^2 = 0.13$].

4. Discussion

The present study has succeeded in demonstrating the presence of a negative transfer effect (or interference) between standard WM and RM training protocols in the water maze. This is often overlooked in the

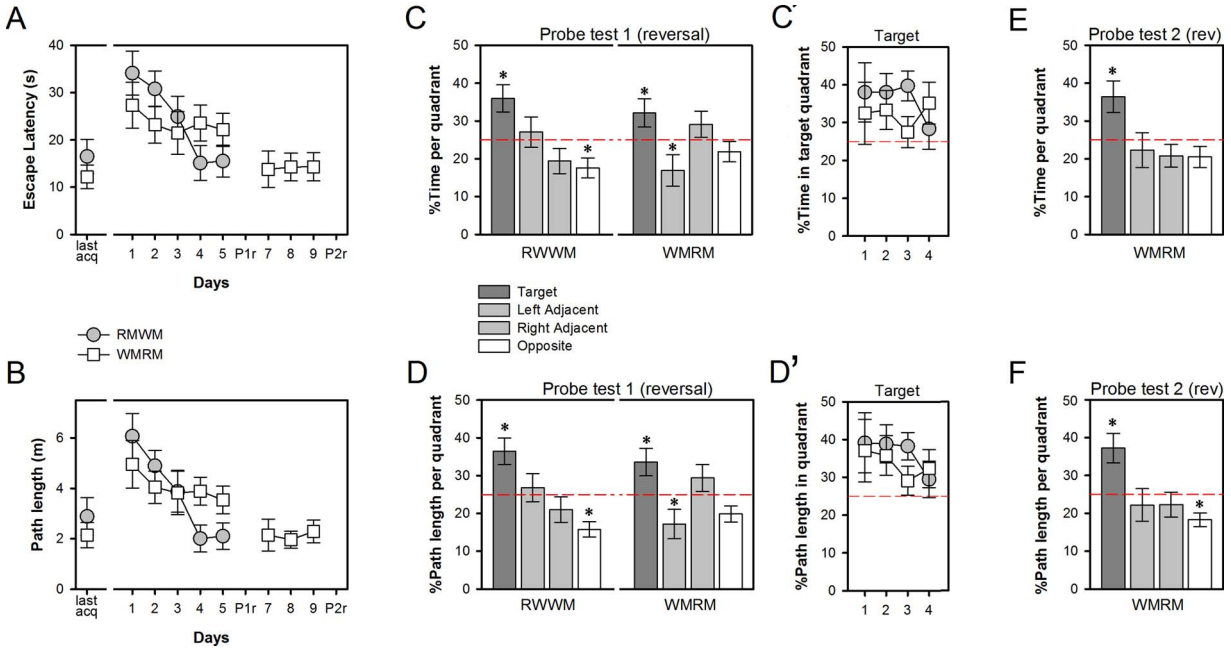

Fig. 6. Performance during reversal learning indexed by escape latency and path length is depicted in (A) and (B), respectively. The performance of the last acquisition day (last acq) is also shown to illustrate the similar impact of switching the platform location. Recall that the RMWM group underwent 5 days of acquisition training, whereas the WMRM group had 10 days of training in total. A probe test was performed 24 h after 5 days of reversal learning. The proportion of time and path length recorded in each of the four quadrants are depicted in (C) and (D), respectively. The evolution across time in the target quadrant was subjected to further analysis (C' and D'). The WMRM group was subjected to 3 additional days of reversal learning afterward, and then another probe test, the performance of which was highly comparable to that achieved by the RMWM group earlier (c.f., C and D). All values express mean ± SEM. Dashed horizontal lines indicate chance level performance. * in (C, C', D, D', E and F) indicates significant deviation from chance level performance based on separate one-sample *t*-tests ($p < 0.05$, two-tailed).

Table 2
Summary of key statistical outcomes indicative of a transfer effect due to prior RM or WM training on the other test. WM performance was captured by the Group by Trials interaction. The transfer effect observed in the initial acquisition phase of RM testing was evidenced by the main effect of Group, and subsequently in the probe test in the form of a *Group* by *Quadrants* interaction. In reversal learning, the transfer effect was observed as a *Group* by *Days* interaction, and subsequently in the probe test that followed reversal learning in the form of a *Group* by *Quadrants* interaction. Not all critical effects and interactions achieved statistical significance. Those achieving significance at $p < 0.05$ levels are highlighted in red. Regardless of statistical significance levels, the observed effect sizes, indexed by (η_p^2), are in the medium-to-large range [33]. The last two columns on the right illustrate the corresponding statistical significance levels (*p*-value) and effect sizes (η_p^2) re-calculated after exclusion of data in day 1 (of RM or WM test) from the respective ANOVAs. This generally improved statistical significance and inflated the effect size. This is also consistent with our expectation that the first day of the second test had not yet revealed the critical procedural difference between RM and WM tests (see Discussion).

Impact of prior RM training on subsequent WM performance: significance and Effect sizes						
Conditions	Statistical effects	Dependent measures	*p*-value	$_p^2$	(*after exclusion of day 1 data*)	
					p-value	$_p^2$
20 s ITI (over 8 days)	Group × Trials	Escape latency	0.024	0.20	0.007	0.28
		Path length	0.098	0.12	0.055	0.15
20 s & 10-min ITI (days 1–10)	Group × Trials	Escape latency	0.049	0.16	0.019	0.22
		Path length	0.083	0.13	0.056	0.15
Impact of prior WM training on subsequent RM performance: significance and Effect sizes						
Conditions	Statistical effects	Dependent measures	*p*-value	$_p^2$	(*after exclusion of day 1 data*)	
					p-value	$_p^2$
Acquisition (days 1–5) days)	Group	Escape latency	0.260	0.06	0.094	0.12
		Path length	0.048	0.16	0.036	0.18
Probe 1 in Acquisition	Group × Quadrants	%Time	0.027	0.12		
		%Path length	0.069	0.10		
Reversal Learning (days 1–5)	Group × Days	Escape latency	0.029	0.11		
		Path length	0.067	0.09		
Probe 1 in Reversal Learning	Group × Quadrants	%Time	0.080	0.09		
		%Path length	0.103	0.09		

behavioural phenotyping literature. Hence, our empirical determination of the magnitude, persistence, and directionality of the transfer effect in C57BL/6 mice – the most common genetic background in the generation of genetic engineered mice – is highly relevant. Our results clearly show that prior training under the WR or RM protocol produced a negative transfer effect on subsequent training in the other task with a medium to large effect size, even when the two tasks were performed in two rooms with distinct distal cues. Although we also showed that extra training could ameliorate the negative transfer effect to a large extent, it might re-emerge later as seen in the early phase of reversal learning (despite protracted acquisition training) in animals having been exposed to the WM training protocol before RM training. Our findings here suggest that caution is warranted over the sequential use of the two water maze protocols in the same animals.

4.1. Direction, magnitude, and persistence of the transfer effects

First, we demonstrated in male C57BL/6 mice that prior exposure to either RM or WM water maze procedures clearly interfered with subsequent performance by the same animals in the other task. Twelve days of training in the first task following minimal procedural and non-spatial pre-training (see Fig. 1) were sufficient to generate a negative effect on performance in the second task in either direction of transfer: RM → WM or WM → RM. Second, with an effect size (η_p^2) approximately 0.1–0.2, the magnitude of the transfer effects was at least medium-to-large according to common convention [33] (see Table 2). The impact of prior RM training (on subsequent WM performance) was only marginally stronger than that of prior WM training (on subsequent RM performance), with obvious overlaps around $\eta_p^2 = 0.16$. Notably statistical significance (at $\alpha = 0.05$) was not always achieved with both the time and distance measures, but the two measures consistently gave rise to a highly similar pattern of outcomes. The current group size was 12 or 13, which corresponds closely to the average sample size per group across the studies reviewed in Table 1. Slightly larger sample sizes are expected to provide the necessary statistical power to increase the likelihood of achieving the desired statistical significance in both performance measures: escape latency and path length. It is also worth noting that we had deliberately adopted a procedure with only 2 trials per day in our RM task, so as to balance the amount of training (trials) between the two tasks, in order to exclude non-specific confounds such as fatigue or familiarity to the first testing room. It is conceivable that the more common 4-trials-per-day RM procedure (see Table 1) would have generated a more substantial transfer effect on subsequent WM performance. This certainly warrants further empirical evaluation and its confirmation would suggest that our estimated effect sizes for the RM → WM negative transfer effect here might underestimate the effects in most published studies summarized in Table 1.

Third, the transfer effects were readily overcome by extra training although they might emerge again later. This was most clearly demonstrated in the WM → RM sequence (Fig. 4). The effect of prior exposure to the WM procedure was obvious within the first 5 days of RM learning, and the probe test 24 h later. Another five days of RM training, however, was sufficient to ameliorate performance to near control levels (i.e., the RMWM group that received spatial RM training first). The “remedial” training for the WMRM group was apparently only effective for the specific RM problem presented during acquisition. When they were subsequently challenged with reversal learning, we saw a similar performance deficit across the first 5 days of reversal learning and in the probe test that followed (see Fig. 5, in comparison to the reversal learning curve by the RMWM group that did not receive any “remedial” acquisition training). The re-emerged performance deficit in the WMRM group was not associated with a stronger reversal effect (comparing between the last acquisition day and the first reversal day), so it is unlikely the direct result of having “remedial” training during the acquisition phase as such. Next, we provided three extra days of “remedial” reversal learning to the WMRM group, and the animals performed at a level attained by the RMWM group 3 days earlier. However, we cannot comment on whether the 5 days and 3 days of “remedial” training in the acquisition and reversal phases received by the WMRM group, respectively, were the minimal required to “normalize” performance. We therefore cannot ascertain if the transfer effect had become more fragile as it re-emerged in the reversal phase, although the effect size appeared to be smaller by comparison with that seen across the first 5 days of acquisition, especially when we examined the respective probe tests (Table 2).

Similarly, the interference by prior RM training on subsequent WM performance could also be overcome by extra training. This is in line with some studies that had extended WM testing up to 21 days when the test followed prior RM testing in the same animals (see e.g., [15]). The WMRM group (as the control group) achieved a statistically significant effect of trials after 5 days [$F_{(1,11)} = 4.94$, $p < 0.05$], while the RMWM took 6 days to yield a significant trials effect [$F_{(1,12)} = 7.21$, $p < 0.05$] despite a slightly larger (by 8%) group size. This may give the impression that the impact of prior RM training was minimal such that one additional day of WM test was sufficient to catch up with control (i.e., WMRM group) performance. This is akin to some studies that had opted to index performance by the amount of training needed to achieve an arbitrarily low escape latency on the second trial [13,21]. However, the effect size (η_p^2) that measures the differential WM performance between groups was 0.20 – amongst the strongest transfer effect we had observed in the present study (Table 2).

It is more difficult to decide if the transfer effect of prior RM training on subsequent WM performance might re-emerge later in our RMWM group. When the delay from trial 1 to trial 2 was extended to 10 min on days 9–10, the RMWM group failed to show any saving in trial 2 while the WMRM group still exhibited appreciable saving (see Fig. 2E and F). The weak performance in the RMWM group here could be triggered by the increase in mnemonic demand or simply the change in the procedure (since the longer delay required placing the animals in a waiting cage between trials 1 and 2). However, we lack solid statistical support that the re-emergence was significant since the critical group by trials interaction did not achieve statistical significance with data restricted to days 9–10.

4.2. Implications & recommendations

Our findings do not question the validity of the two tests examined here as such, but draw attention to the presence of an appreciable experimental confound whenever the RM and WM tests are performed consecutively in the same animals, in whichever order. Whatever the second outcome may be, it must be considered against the background of a negative transfer effect induced by the first test. This concern has received limited discussion in the literature. We have only identified one mouse study that had explicitly counterbalanced the order of tests [22]. Here, we showed that the prevalent RM → WM sequence is as vulnerable as the alternative WM → RM sequence. As shown in Table 1, there are many studies with a RMWM design reporting unusual difficulty in the acquisition of reliable WM performance in mutants as well as control mice. Many of them required extra training to yield sufficiently clear statistical evidence of learning (see Introduction). Whenever a specific WM deficit in mutants is reported following normal acquisition of RM, it may stem from a stronger negative transfer effect (e.g., [13]) especially when the controls took some days to establish appreciable WM performance [9]. We hasten to caution that the presence of a deficit in both RM and WM tests (e.g., [15]) may not imply that the second deficit would be relatively free from contamination of a negative transfer effect. Correlative analyses of our data did not suggest that better performance in the first test was associated with poorer performance in the second test. So, at least in our mice, it was the exposure to the first test, rather than performance levels in the test as such, that mattered. Hence, important conclusions relying on results obtained in the second test (be it RM or WM) ought to be replicated under conditions free from the associated

transfer effects. This could involve the use of other appetitive spatial learning paradigms, such as radial arm maze, radial water maze, Barnes maze or dry land version of the water maze (e.g., [29,34–42]). In future experiments, water maze RM and WM tests ought to be conducted in separate cohorts of mice whenever possible. Otherwise, it is imperative to counterbalance the order of tests, so as to gauge any impact of potential transfer effects. At the same time, the option to report the two tests as obtained in two separate cohorts is possible if the experimenter so desires subsequently against reporting the second test altogether – perhaps to avoid addressing phenotypes that only emerged when the RM or WM test was conducted as the second test. However, such an approach in effect would half the originally intended sample size. Thus, it should only be considered when the final sample sizes (comprising only water maze naïve animals) are still sufficient to support sensible interpretation of any null effect.

4.3. The psychological nature of the negative transfer effects

As explained in the Introduction, one view is that conflicting strategies adopted during RM and WM training retarded the acquisition of the other task. The RM and WM tasks are driven by the same motivation to escape from the water and have identical physical demands. The obvious critical difference is whether the location of the escape platform was relatively stable (in RM) or variable (in WM) between days. This defining feature, however, cannot be detected until the second day of the second task. It follows that the negative transfer effect should emerge, at the earliest, on the second day of the second test but not on the first day (because when considered alone, the testing procedures of RM and WM were effectively identical). This is largely consistent with our data. Indeed, all critical terms in the ANOVAs suggestive of a negative transfer effect (viz., the group effect in the RM test, and the group by trials interaction in the WM test) achieved higher levels of statistical significance when excluding data of the first day from the analyses (Table 2). As shown in Fig. 3A–B, the poorer acquisition performance of the WMRM group relative to the RMWM group emerged over days, and there was no evidence for a negative transfer effect in the first day of RM training. It was the overall acquisition performance across days and the first probe test that most readily distinguished the two groups. Similarly, we did not see a substantial difference between groups in the performance in the WM task as indexed by the saving from trial 1 to trial 2 on the first day (discernible in Fig. 2A–B). Instead, WM performance by the RMWM group was less robust across days compared with the WMRM group.

4.4. Some caveats

Although we showed that RM training (with the never-changing platform location) could undermine the use of flexible memory demanded by the daily switching of platform location during WM test, we cannot confirm here that the effect was attributable to a strategy that had biased the animals to return to yesterday's platform location during subsequent WM training. This is expected to play a role when we consider reports in rats demonstrating that a similar delayed matching-to-place WM protocol could bias the rats to return to yesterday's platform location in the first trial (performed as a probe test) on the next day [25,26]. If such a proactive interference effect had contributed significantly to the deficient WM performance observed in the RMWM group, a performance deficit ought to be visible on trial 1 when the proactive interference (hence, bias for yesterday's platform location) was at its maximum. However, this predicted effect was only apparent when the ITI was extended to 10 min (right panel of Figs. 3C–D) – notably on the last day of test (Fig. 3A and B). By contrast, trial 1 performance hardly differed between the two groups when the ITI was limited to 20 s (left panel of Figs. 3C–D). At this minimal ITI, the WM deficiency in the RMWM group relative to the WMRM group stemmed solely from poorer trial 2 performance. If we had conducted the probe tests on trial 1 as Steele & Morris [25] and McGarrity et al. [26] did in rats (for assessing the bias of yesterday's platform location), rather than on trial 2 (designed to look for a bias towards platform location revealed on trial 1), a more direct comparison could have been made. It is worth noting that these demonstrations of proactive interference in rats [25,26] had invariably employed a 4-trial per day protocol, which may be expected to instil a stronger memory trace of yesterday's target location and/or a higher likelihood of generalizing the win-stay strategy (acquired within-days) across days. Unfortunately, we have not identified any similar rat studies with a 2-trial per day protocol, so we cannot exclude the possibility of a species difference.

Similarly, we also cannot confirm that animals with prior WM experience were disadvantaged in the subsequent RM test because they failed to carry over information acquired on previous days. If so, one would also expect that the deficit in the WMRM group seen across the first five days of RM acquisition to be more prominent in the first trial. Yet, the group by trials interaction was far from significance [F's < 1]. Analysis of a saving score that contrasts the first trial of a given day with the second trial of the previous day also did not reveal any group difference (data not show). So it seems that neither of these accounts can satisfactorily explain our observations here; and the precise psychological account of the negative transfer effects between water maze RM and WM procedures remain to be elucidated. Any satisfactory account should also resolve why a win-stay strategy that is apparently common to RM and WM tests does not instead result in a positive transfer effect.

4.5. Other sources of interference

Although we may not pinpoint exactly the source of interference responsible for the negative transfer effects demonstrated here, our experimental design has permitted us to exclude some alternatives. First of all, we may exclude that memory of the platform location(s) as such resulted in the negative transfer effects. If so, the RM → WM sequence ought to be more interfering than WM → RM as the WM training did not favour the development of persistent preference for any one specific location (as evidenced by the probe tests in Fig. 3). By conducting the RM and WM tests in two distinct rooms, we precluded the possibility that preference for any specific location(s) defined by distal cues in one room can interfere in the other room. We have previously shown that the room switching used here was sufficient to eliminate the reversal effect in mice of C57BL/6 background [24] in a RM procedure whereby the reversal phase was replaced by *new* RM learning in another room. Incidentally this demonstration by Singer et al. [24] further showed that a RM_{room1} → RM_{room2} sequence was not associated with any negative transfer effect in the second RM task. We may therefore exclude the possibility that the negative transfer effects (WM → RM as well as RM → WM) here stemmed solely from the requirement to learn the second task in a new testing room. Interference may arise from exposure to a new set of distal cues and/or suppression of the previously acquired map. It may be argued that the additional cognitive demands present in the second task could be a parsimonious explanation for poorer learning in the second test. This would predict that a RM_{room1} → RM_{room2} or WM_{room1} → WM_{room2} transfer should be similarly effective. Singer et al. [23], however, had provided grounds to reject the former prediction even though the latter (WM_{room1} → WM_{room2} transfer) has not been tested.

4.6. Limitations

First, we did not examine parameters that may modulate the size of the transfer effects, such as the time interval between tests. Our ability to detect significant negative transfer effects may partly be attributed to our decision to minimize this interval. A systematic analysis of potential parameters, including training duration and number of trials per day could shed light on how one may effectively reduce or enhance the transfer effects. Second, we did not counterbalance the two rooms here

so as to allow opportunities for RM and WM to be conducted in each room. Finally, this study only included adult male subjects of one mouse strain. Although the widespread use of the C57BL/6 strain maximizes the implications of our findings, especially in the behavioural analysis of genetically engineered mice, the generality of our findings across species, strains, and sexes clearly warrants further empirical investigation.

Conflicts of interest

None.

Acknowledgements

This study was supported in part by funding from the ETH Zurich. SD and PS were recipients of a studentship from the Neural Plasticity & Repair – National Centre for Competence in Research (NCCR) funded by the Swiss National Science Foundation. We wish to express our gratitude to the excellent technical services and supports provided by Mr Peter Schmid. The additional support from the Hong Kong Polytechnic University (grant ZE64) to BKY during the preparation of this article is also acknowledged.

References

[1] R.G. Morris, P. Garrud, J.N. Rawlins, J. O'Keefe, Place navigation impaired in rats with hippocampal lesions, Nature 297 (5868) (1982) 681–683.
[2] C.V. Vorhees, M.T. Williams, Morris water maze: procedures for assessing spatial and related forms of learning and memory, Nat. Protoc. 1 (2) (2006) 848–858.
[3] A.V. Terry, Spatial navigation (Water maze) tasks, in: J.J. Buccafusco (Ed.), Methods of Behavior Analysis in Neuroscience, 2nd ed., CRC Press/Taylor & Francis, 2009.
[4] R. Morris, Developments of a water-maze procedure for studying spatial learning in the rat, J. Neurosci. Methods 11 (1984) 47–60.
[5] H. Hodges, P. Sowinski, J.D. Sinden, C.A. Netto, A. Fletcher, The selective 5-HT3 receptor antagonist, WAY100289, enhances spatial memory in rats with ibotenate lesions of the forebrain cholinergic projection system, Psychopharmacology (Berl) 117 (3) (1995) 318–332.
[6] J. Nunez, Morris water maze experiment, J. Visualized Exp. 19 (2008) e897.
[7] M.C. Buhot, M. Wolff, M. Savova, G. Malleret, R. Hen, L. Segu, Protective effect of 5-HT1B receptor gene deletion on the age-related decline in spatial learning abilities in mice, Behav. Brain Res. 142 (1–2) (2003) 135–142.
[8] L. Giménez-Llort, S.A. Masino, L. Diao, A. Fernández-Teruel, A. Tobeña, L. Halldner, et al., Mice lacking the adenosine A1 receptor have normal spatial learning and plasticity in the CA1 region of the hippocampus, but they habituate more slowly, Synapse 57 (1) (2005) 8–16.
[9] C.Z. Liu, J.C. Yu, H.Y. Cheng, Z.G. Jiang, T. Li, X.Z. Zhang, et al., Spatial memory performance and hippocampal neuron number in osteoporotic SAMP6 mice, Exp. Neurol. 201 (2) (2006) 452–460.
[10] A. Bour, J. Grootendorst, E. Vogel, C. Kelche, J.C. Dodart, K. Bales, et al., Middle-aged human apoE4 targeted-replacement mice show retention deficits on a wide range of spatial memory tasks, Behav. Brain Res. 193 (2) (2008) 174–182.
[11] L.H. Torres, R.C. Garcia, A.M. Blois, L.M. Dati, A.C. Durão, A.S. Alves, et al., Exposure of neonatal mice to tobacco smoke disturbs synaptic proteins and spatial learning and memory from late infancy to early adulthood, PLoS One 10 (8) (2015) e0136399.
[12] G. Malleret, R. Hen, J.L. Guillou, L. Segu, M.C. Buhot, 5-HT1B receptor knock-out mice exhibit increased exploratory activity and enhanced spatial memory performance in the Morris water maze, J. Neurosci. 19 (14) (1999) 6157–6168.
[13] H. Zeng, S. Chattarji, M. Barbarosie, L. Rondi-Reig, B.D. Philpot, T. Miyakawa, et al., Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory, Cell 107 (5) (2001) 617–629.
[14] X.G. Huang, B.K. Yee, S. Nag, S.T. Chan, F. Tang, Behavioral and neurochemical characterization of transgenic mice carrying the human presenilin-1 gene with or without the leucine-to-proline mutation at codon 235, Exp. Neurol. 183 (2) (2003) 673–681.
[15] C. Janus, Search strategies used by APP transgenic mice during navigation in the Morris water maze, Learn. Mem. 11 (3) (2004) 337–346.
[16] S.J. Zhou, M.E. Zhu, D. Shu, X.P. Du, X.H. Song, X.T. Wang, et al., Preferential enhancement of working memory in mice lacking adenosine A(2A) receptors, Brain Res. 1303 (2009) 74–83.
[17] J. Espallergues, L. Galvan, F. Sabatier, V. Rana-Poussine, T. Maurice, A. Chatonnet, Behavioral phenotyping of heterozygous acetylcholinesterase knockout (AChE +/−) mice showed no memory enhancement but hyposensitivity to amnesic drugs, Behav. Brain Res. 206 (2) (2010) 263–273.
[18] G. D'Agostino, R. Russo, C. Avagliano, C. Cristiano, R. Meli, A. Calignano, Palmitoylethanolamide protects against the amyloid-β25-35-induced learning and memory impairment in mice, an experimental model of Alzheimer disease, Neuropsychopharmacology 37 (7) (2012) 1784–1792.
[19] P. Liu, Z. Huang, N. Gu, Exposure to silver nanoparticles does not affect cognitive outcome or hippocampal neurogenesis in adult mice, Ecotoxicol. Environ. Saf. 87 (2013) 124–130.
[20] Z. Xu, J. Li, F. Zhang, Y. Wu, Y. Gao, J. Liang, et al., Working memory task decreases the survival of newly born neurons in hippocampus, Neurobiol. Learn. Mem. 95 (3) (2011) 239–247.
[21] N.Z. Rahman, S.M. Greenwood, R.R. Brett, K. Tossell, M.A. Ungless, R. Plevin, et al., Mitogen-activated protein kinase phosphatase-2 deletion impairs synaptic plasticity and hippocampal-dependent memory, J. Neurosci. 36 (8) (2016) 2348–2354.
[22] C.J. Lawson, J. Homewood, A.J. Taylor, The Effects of L-glucose on memory in mice are modulated by peripherally acting cholinergic drugs, Neurobiol. Learn. Mem. 77 (1) (2002) 17–28.
[23] P. Singer, D. Boison, H. Möhler, J. Feldon, B.K. Yee, Deletion of glycine transporter 1 (GlyT1) in forebrain neurons facilitates reversal learning: enhanced cognitive adaptability? Behav. Neurosci. 123 (5) (2009) 1012–1027.
[24] P. Singer, B.K. Yee, J. Feldon, T. Iwasato, S. Itohara, T. Grampp, et al., Altered mnemonic functions and resistance to *N*-methyl-D-Aspartate receptor antagonism by forebrain conditional knockout of glycine transporter 1, Neuroscience 161 (2) (2009) 635–654.
[25] R.J. Steele, R.G. Morris, Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5, Hippocampus 9 (2) (1999) 118–136.
[26] S. McGarrity, R. Mason, K.C. Fone, M. Pezze, T. Bast, Hippocampal neural disinhibition causes attentional and memory deficits, Cereb. Cortex 27 (9) (2017) 4447–4462.
[27] S.L. Inman-Wood, M.T. Williams, L.L. Morford, C.V. Vorhees, Effects of prenatal cocaine on Morris and Barnes maze tests of spatial learning and memory in the offspring of C57BL/6J mice, Neurotoxicol. Teratol. 22 (4) (2000) 547–557.
[28] P. Singer, J. Hauser, L. Llano Lopez, D. Peleg-Raibstein, J. Feldon, P.A. Gargiulo, B.K. Yee, Prepulse inhibition predicts working memory performance whilst startle habituation predicts spatial reference memory retention in C57BL/6 mice, Behav. Brain Res. 242 (2013) 166–177.
[29] S. Dubroqua, L. Serrano, D. Boison, J. Feldon, P.A. Gargiulo, B.K. Yee, Intact working memory in the absence of forebrain neuronal glycine transporter 1, Behav. Brain Res. 230 (1) (2012) 208–214.
[30] S. Pietropaolo, J. Feldon, B.K. Yee, Age-dependent phenotypic characteristics of a triple transgenic mouse model of Alzheimer disease, Behav. Neurosci. 122 (4) (2008) 733–747.
[31] T. Bast, I.A. Wilson, M.P. Witter, R.G. Morris, From rapid place learning to behavioral performance: a key role for the intermediate hippocampus, PLoS Biol. 7 (4) (2009) e1000089 21.
[32] M. Pezze, T. Bast, Dopaminergic modulation of hippocampus-dependent learning: blockade of hippocampal D1-class receptors during learning impairs 1-trial place memory at a 30-min retention delay, Neuropharmacology 63 (4) (2012) 710–718.
[33] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Erlbaum, Hillsdale, NJ, 1988.
[34] D.S. Olton, R.J. Samuelson, Remembrance of places passed: spatial memory in rats, J. Exp. Psych. Anim. Behav. Processes 2 (2) (1976) 97–116.
[35] C.A. Barnes, Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat, J. Comp. Physiol. Psychol. 93 (1) (1979) 74–104.
[36] A.J. Fader, P.E. Johnson, G.P. Dohanich, Estrogen improves working but not reference memory and prevents amnestic effects of scopolamine of a radial-arm maze, Pharmacol. Biochem. Behav. 62 (4) (1999) 711–717.
[37] J.E. Gresack, K.M. Frick, Male mice exhibit better spatial working and reference memory than females in a water-escape radial arm maze task, Brain Res. 982 (1) (2003) 98–107.
[38] J.N. Crawley, What's Wrong with My Mouse? Behavioral Phenotyping of Transgenic and Knockout Mice, 2nd ed., John Wiley & Sons, Inc, Hoboken, New Jersey, 2007.
[39] L. Llano Lopez, J. Hauser, J. Feldon, P.A. Gargiulo, B.K. Yee, Evaluating spatial memory function in mice: a within-subjects comparison between the water maze test and its adaptation to dry land, Behav. Brain Res. 209 (1) (2010) 85–92.
[40] S.C. Penley, C.M. Gaudet, S.W. Threlkeld, Use of an eight-arm radial water maze to assess working and reference memory following neonatal brain injury, J. Visualized Exp. 82 (2013) 50940.
[41] J.M. Fine, D.B. Renner, A.C. Forsberg, R.A. Cameron, B.T. Galick, C. Le, et al., Intranasal deferoxamine engages multiple pathways to decrease memory loss in the APP/PS1 model of amyloid accumulation, Neurosci. Lett. 584 (2015) 362–367.
[42] E.J. Melief, E. Cudaback, N.L. Jorstad, E. Sherfield, N. Postupna, A. Wilson, et al., Partial depletion of striatal dopamine enhances penetrance of cognitive deficits in a transgenic mouse model of Alzheimer's disease, J. Neurosci. Res. 93 (9) (2015) 1413–1422.