The magic number 70 (plus or minus 20): Variables determining performance in the Rodent Odor Span Task

L. Brooke April, Katherine Bruce, Mark Galizio *

University of North Carolina Wilmington, United States

ARTICLE INFO

ABSTRACT

The olfactory span task (OST) uses an incrementing non-matching to sample procedure such that the number of stimuli to remember increases during the session. The number of consecutive correct responses (span length) and percent correct as a function of the memory load have been viewed as defining rodent working memory capacity limitations in several studies using the OST. However, the procedural parameters of the OST vary across experiments and their effects are not well understood. For example, in several studies, the number of stimuli to remember is confounded with the number of comparison stimuli displayed in the test arena. Experiment 1 addressed whether performance is influenced by the number of comparison choices available on any given trial (2, 5, 10) as well as the number of odor stimuli to remember during a session (12, 24, 36). Performance was most accurate when the number of stimuli to remember was low, as would be expected from a working memory interpretation of OST. However, accuracy was also affected by the number of comparison stimulus choices. High levels of accuracy were seen even with 36 odors, suggesting that the capacity for odor memory in rats was greater than suggested by previous research. Experiment 2 attempted to define this capacity by programming sessions with 36, 48 or 72 stimuli to remember in a group of rats that had previously received extensive OST training. Highly accurate performance (80% correct or better) was sustained throughout the session at even the greatest memory loads, arguing strongly against the notion that the OST models the limited capacity of human working memory. Experiment 3 explored the possibility that stimulus control in the OST is based on relative stimulus familiarity, rather than recognition of stimuli not yet presented during the current session. Number of odor cups visited increased with the number of comparisons in the arena, but rats rarely sampled all of the comparison odors before responding. However, on probe trials which included only stimuli that had been presented during the session, latency to respond and number of comparisons sampled was sharply increased. These data suggest that responding in the OST is determined not just by relative familiarity, but rather by a more specific "what-when" or perhaps "how long ago" form of stimulus control.

© 2013 Elsevier Inc. All rights reserved.

A number of procedures have been used to measure memory or retention of stimulus control in non-humans, including delayed alternation, the radial arm maze, the Morris swim task, the object preference task and a variety of delayed match- and non-match-to-sample tasks (see Alvarado & Bachevalier, 2008; Dudchenko, 2004, for reviews). Although these tasks are often referred to as assessments of "working memory", it should be understood that this term is used somewhat differently than in the human cognitive literature. Generally, tasks are defined as measuring working memory in non-humans if they involve: "a short-term memory for an object, stimulus or location that is used within a testing session, but not typically

* Corresponding author at: Department of Psychology, University of North Carolina Wilmington, Wilmington, NC 28403, United States.
E-mail address: Galizio@uncw.edu (M. Galizio).

between sessions" (Dudchenko, 2004, p. 700). For example, in a delayed match-to-sample procedure, a sample stimulus is presented and, following some type of observing response, is terminated. After a delay interval programmed by the experimenter, two or more comparison stimuli are presented, one of which is identical to the sample. Responding to the stimulus that "matches the sample" indicates that stimulus control has persisted across the delay interval. Generally, in such studies, a delay gradient is obtained such that the longer the interval between the presentation of the sample and the comparison array, the poorer the performance; this is often referred to as a forgetting function (Wright, 2007).

However, as Wright (2007) has pointed out, most of the procedures used to study working memory in non-humans focus on the effects of a delay interval on retention of a single item and, as such, may have limited validity with respect to the way the construct is defined in humans. In human research, working memory is generally viewed as *short lived*, affected by the delay interval, but also as of *limited capacity*, affected by the number of items to be remembered (Gathercole, 2009). In contrast, there are few studies that explore capacity limitations using the standard working memory tasks in animals, and those have not found severe capacity limitations. For example, radial arm maze performance in rats remains highly accurate with up to 24 different arm locations (Cole & Chappell-Stephenson, 2003; Roberts, 1979). However, recently the odor span task (OST) developed by Dudchenko, Wood and Eichenbaum (2000) for rodents has become a popular technique to study working memory under conditions in which the number of stimuli to remember can be manipulated.

Dudchenko et al. (2000) trained rats to dig in cups filled with scented sand to obtain food reinforcement in an arena with spaces for cups to be placed about the perimeter. On the first trial, a single cup (Odor 1) is placed in the arena and the rat is permitted to dig in the cup to obtain food. On Trial 2, two cups are randomly placed in the arena: one containing Odor 1 (unbaited) and the other a new scent (Odor 2). Responses to Odor 2, but not Odor 1, are reinforced, so the procedure follows the form of a non-match-to-sample task. On the next trial, three stimuli are presented: Odors 1 and 2 (unbaited) and a new odor, Odor 3 (baited). The procedure continues in this way throughout the session with one new odor added on each trial; thus, the task could be described as an incrementing non-match-to-sample procedure—with the number of sample stimuli to remember increasing as the session progresses. The main dependent measure used in this study was the span length: the number of consecutive correct responses minus one (because the subject had only one choice on the first trial of each span). Dudchenko et al. found mean span lengths of just over eight odors for the rats tested. Several additional researchers have used variations on the OST in rats and mice and have consistently found span lengths to range between 8 and 10 odors (e.g., MacQueen, Bullard, & Galizio, 2011; Young, Kerr et al., 2007).

The OST is increasingly being used as an assay to explore the effects of neurobiological manipulations on working memory capacity. For example, Turchi and Sarter (2000) showed that rats' performance on the task was impaired by lesions of the basal forebrain cholinergic system. Young and colleagues have found deficits in span length and accuracy in human amyloid over-expressing mice and in alpha7-nicotinic acetylcholine receptor knockout mice, both models of cognitive impairment (Young, Crawford et al., 2007; Young, Sharkey, & Finlayson, 2008); further, Cui et al. (2011) showed enhanced OST performance in NMDA NR2B subunit over-expressing mice. Behavioral pharmacological analyses have shown that the OST is sensitive to drug effects as well. For example, nicotine appears to enhance memory capacity (Rushforth, Allison, Wonnacott, & Shoaib, 2010; Rushforth, Steckler, & Shoaib, 2011) and NMDA antagonists impair it (Galizio, Deal, Hawkey, & April, 2013; MacQueen et al., 2011; Rushforth et al., 2011).

Most of the studies described above focused on span length as the index of memory capacity. However, Dudchenko et al. (2000) also examined accuracy as a function of the number of stimuli to remember (number of sample stimuli presented) and found a decrease in accuracy with increasing memory load. Although this finding supports the validity of the OST as a measure of working memory capacity, there exists a confound between the number of stimuli to remember and the number of comparison stimuli from which to choose. We have attempted to separate these variables by limiting the number of stimuli in the arena to five even as the number of new stimuli presented continued to increment (Galizio et al., 2013; MacQueen et al., 2011). Thus, each trial after the fourth trial included one new stimulus (S+) as well as four previously presented stimuli (S−) randomly chosen from the pool of odors presented in the previous trials of the session. We were able to replicate the finding of decreased accuracy as a function of the number of stimuli to remember, even with the number of comparison stimuli restricted. However, in both studies we also observed a relatively sharp decrease in accuracy during the first few trials (when the number of comparison stimuli was increasing along with the memory load) and a very shallow decrease in accuracy after the number of comparison stimuli had reached the maximum number of five. Indeed, accuracy remained well above chance with up to 24 stimuli in these studies (Galizio et al., 2013; MacQueen et al., 2011). These findings raise questions about the proper interpretation of the OST. Is performance determined exclusively by the number of stimuli to remember (as is typically assumed), or does the number of comparison stimuli also affect OST performance? The purpose of Experiment 1 was to assess the relative contribution of these two variables in the OST.

Experiment 1: Effects of number of comparison stimuli and number of stimuli to remember on OST performance

Method

Subjects

Six male Holtzman (Sprague–Dawley) albino rats 90–120 days old at the beginning of the experiment served as subjects. Water was available ad lib., but access to food was restricted to maintain approximately 85% of free feeding weight. Subjects were individually housed and maintained on a 12:12 hr light/dark cycle.
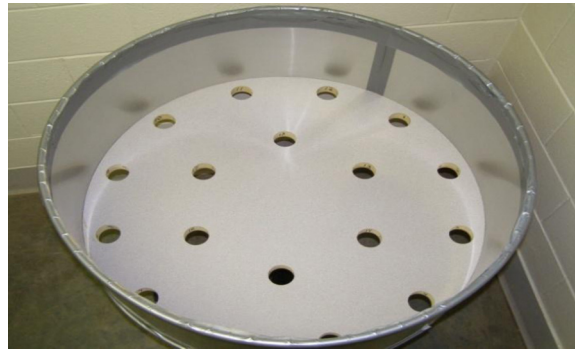
Fig. 1. Arena apparatus used for all experiments.

Apparatus

The apparatus was a circular open-field arena (94 cm diameter) with 18 drilled holes (5 cm in diameter) arranged in two circular arrays on the floor (described previously in Galizio et al., 2013; MacQueen et al., 2011; also see Fig. 1). Plastic stimulus cups (2 oz) were placed in the holes. Aluminum baffling approximately 30 cm high surrounded the perimeter, and a white masking noise (70 dB) was presented throughout the session. The arena apparatus and a separate holding cage (20 cm × 30 cm) located on a table adjacent to the arena were housed in a small room with a video camcorder positioned in the ceiling so that each session could be digitally recorded.

Stimuli

Odor stimuli were presented by covering the plastic cups with opaque plastic lids that had been scented by storing them in plastic containers containing odorants such as aromatic oils and household spices. A total of 36 difference odorants were used, and all were purchased from Great American Spice Co. (see Table 1 for complete list of odor stimuli).

Procedure

Pre-training. First exposure to the arena included one or more cups containing only sugar pellets (45 mg Bio Serv). When the rat was readily consuming the pellets, trials were conducted with baited cups partially covered by an unscented plastic lid. As rats became successful at removing the lid and retrieving pellets, the lid was gradually positioned such that it covered the cup completely. Once lid removal was reliable, the non-matching-to-sample (Odor Span) training began. For the remainder of the experiment, a response was defined as the displacement of a lid by 1 cm or more by the front paws or snout.

OST training. On these sessions, plastic lids were scented with one of the 36 odors. A single scented lid and baited cup was randomly placed in one location with the remaining 17 holes filled with empty cups. Animals were placed in the arena and remained there until removal of the lid from the stimulus cup and retrieval of the sucrose pellet. After an inter-trial interval (ITI) of approximately 30 s (spent in a holding cage in the same room), the rat was returned to the arena now with two stimulus cups placed in random locations (Trial 2). For Trial 2, one cup was covered with a lid of the same odor presented on the previous trial (now unbaited, S−) while the second was covered with a differently scented lid (baited, S+). If a response to the new (non-matching) odor occurred, the next trial included three stimuli in the arena: two cups with lids scented

Table 1
Odor list for Experiments 1 (36 odors, top panel) and 2 (36 odors, bottom panel).

Experiment 1					
Allspice	Almond	Anise	Bay	Beet	Caraway
Carob	Celery	Cherry	Cinnamon	Cloves	Coriander
Cumin	Dill	Fennel	Fenugreek	Garlic	Ginger
Grape	Lime	Marjoram	Mustard	Nutmeg	Onion
Orange	Oregano	Paprika	Peach	Rosemary	Sage
Savory	Spinach	Sumac	Thyme	Turmeric	Vanilla
Experiment 2					
Annatto	Apple	Apricot	Banana	Bergamot	Black Walnut
Blackberry	Blueberry	Brandy	Bubble Gum	Butter	Butterscotch
Champagne	Chocolate	Coconut	Coffee	Hickory	Honey
Lemon	Licorice	Maple	Peanut Butter	Pecan	Peppermint
Pineapple	Pistachio	Pumpkin	Raspberry	Rootbeer	Rum
Sandalwood	Sassafras	Strawberry	Tangerine	Watermelon	Worcestershire

with the previously presented odors and not baited with a food pellet (S−), and a third cup baited and covered with a lid scented with a new odor (S+). This sequence of adding one new baited stimulus to the pool of previously presented odors (which were no longer baited) continued throughout the session until an error was made. Errors were defined as a response made to a previously presented odor. A correction procedure was used such that after an error occurred, the trial continued until the subject responded to the correct stimulus and retrieved the food pellet. If a correct response was not made within 1 min, the trial was terminated and the rat was placed with its snout near the correct lid to assure that it contacted the new scent. Following an error, the incrementing procedure was reset such that the next trial presented a single new odor with no comparison stimuli, and the number of stimuli incremented as before following subsequent correct responses. Sessions terminated after completion of 24 trials, and animals were advanced to the Parametric Span training phase when either of the following mastery criteria was met: (1) making ten or more consecutive correct responses within a session or (2) making five or more consecutive correct responses for two consecutive sessions. Between four and 12 sessions were required to meet criterion (mean = 8.0 sessions).

Parametric Span Task. Two independent variables were manipulated during this phase: the number of comparison choices (2, 5, or 10) and number of sample stimuli to remember (12, 24, 36). The conditions formed by this 3×3 factorial design were studied within-subject, with one of the nine cells of the factorial in effect per session. Thus, in any given session, the number of comparison stimuli in the arena was permitted to increment to 2, 5, or 10, and the number of sample stimuli to remember was 12, 24, or 36. The order of conditions was determined randomly without replacement until each subject had completed each of the nine conditions (one cycle). Then two additional training cycles of sessions were conducted such that at the end of the experiment, each animal had been exposed to three determinations of each of the possible combinations of conditions (resulting in a $3 \times 3 \times 3$ factorial for statistical analysis). The procedure for parametric span differed from initial OST training in that the session progressed (the number of stimuli to remember continued to increment) without resetting following an error. For 2-choice sessions, the number of comparison stimuli incremented to two on Trial 2 (one S+, one S−), and the third and all subsequent trials included two comparison stimuli; one cup with a previously presented odor (selected randomly) and not baited with a food pellet (S−) and one cup baited and covered with a lid scented with a new odor (S+). For 5-choice sessions, the number of comparison choices incremented on each subsequent trial up to Trial 5; then all later trials contained five stimuli: one new odor stimulus (S+) and four familiar, previously presented odor stimuli (S−). For 10-choice sessions the number of comparisons incremented up to Trial 10, and all subsequent trials contained one S+ and nine S− stimuli. The number of different odors presented was 12, 24, or 36, and session duration thus was either 12, 24, or 36 trials, depending on the condition.

Pellet detection and scent marking controls. Non-baited control trials were conducted weekly to ensure behavior was guided by the odor of the lid and not the odor of the food pellet. During a designated pellet detection control session, six trials in which no food pellets were present in the S+ stimulus cup were semi-randomly distributed throughout the session. On these trials, the pellet was delivered into the cup by hand once a response to the S+ was made. Performances on baited trials during these sessions were not significantly more accurate than non-baited trials (mean baited trials = 84.4%, mean unbaited trials = 84.1%, $p > .05$), indicating that performance was not guided by the scent of the pellet. To ensure that behavior could not be guided by scent marks on the lids, lids were changed after each trial and were not reused within that session.

Results and discussion

A $3 \times 3 \times 3$ factorial ANOVA was conducted on accuracy (percent correct) revealing significant main effects for number of sample odors [$F(2,10) = 31.37$, $p < .05$], number of comparison stimuli [$F(2,10) = 24.15$, $p < .05$] and training cycles [$F(2,10) = 11.49$, $p < .05$], and no significant interactions. Overall accuracy increased with training cycle going from a mean percent correct of 71.8 on Cycle 1 to 81.3 and 83.4 on Cycles 2 and 3 respectively. Post hoc tests (LSD) showed that Cycles 2 and 3 differed significantly from Cycle 1, but not from one another ($p < .05$). In the absence of any significant interactions, subsequent data presentations and analyses highlight number of sample odors and comparisons collapsed across the training cycles.

Fig. 2 shows the mean percent correct across these conditions (top panel) and reveals that accuracy was inversely related to both independent variables. Specifically, accuracy was highest when there were 12 sample odors (black circles) and declined significantly with 24 and 36 ($p < .05$), but at 24 and 36 (white squares and black triangles, respectively) accuracies did not differ significantly from one another. Percent correct was highest when there were two comparisons to choose between, and declined as the number of comparisons increased. Each comparison condition differed significantly from the other two conditions ($p < .05$). In sum, although the number of comparison stimuli clearly affected performance in the OST, there was an independent effect of the number of sample stimuli to remember, at least with respect to overall percent correct.

Most OST studies have focused on span length as the main measure of working memory and this is shown in bottom panel of Fig. 2. Span length was inversely related to number of comparison stimuli with longer span lengths (mean = 7.8 items) occurring when only two comparisons were present and shorter span lengths when there were 5 (mean = 4.9 items) or 10 (mean = 3.8 items) comparisons [$F(2,10) = 32.38$, $p < .05$]. That increasing the number of comparison stimuli reduced span length raises important concerns regarding the use of this measure as an index of capacity effects in previous studies (Rushforth et al., 2010; Young et al., 2008). Our finding suggests that the previously reported results may have been
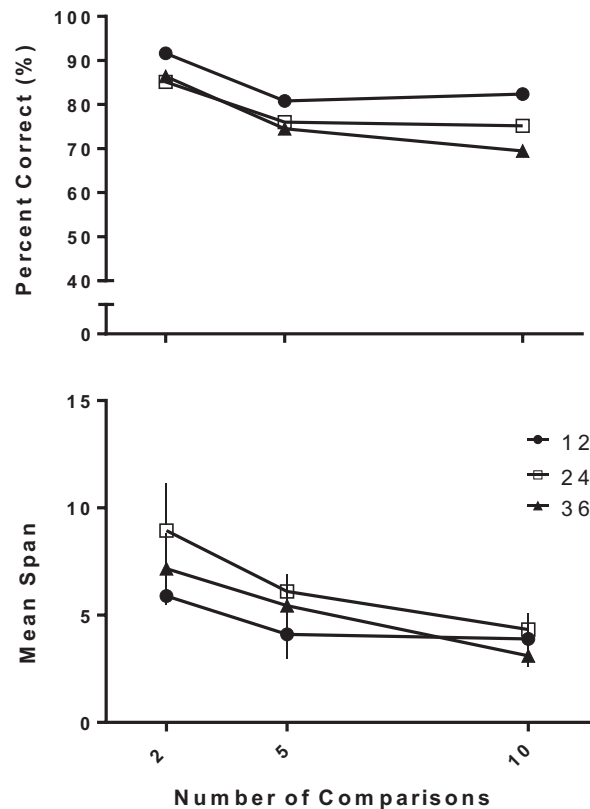
Fig. 2. Experiment 1. Mean accuracy (percent correct; depicted in top panel) and mean span (bottom panel) across the three stimulus conditions (12, 24, 36) and as a function of number of comparison choices available (2, 5, 10). Black circles represent 12 stimulus conditions, white squares are 24 stimuli, and black triangles are 36 stimuli. Error bars represent SEM (in some cases the error bar is obscured by the data point symbol).

determined as much by the increasing number of comparison stimuli as by the increasing memory load. Finally, number of samples was not significantly related to span length [$F(2,10) = 1.94$, $p > .05$], and the sample number × comparison number interaction was not significant [$F < 1$]. Also of interest is the contrast between the relatively low span lengths and the high overall accuracies; even after making their initial error (which defines the span length), animals generally continued to perform at above chance levels of accuracy for the remainder of the session. Consistent with this point, although accuracy was highest in the 12 sample odor condition, performance was well above chance with 24 and 36 sample stimuli as well.

In order to more directly examine the relationship between memory load and accuracy within a session, Fig. 3 shows percent correct as a function of the number of sample stimuli to remember plotted in bins of three trials. The top panel of Fig. 3 shows performance during the 12 sample stimulus sessions. Very little change in accuracy within session was evident as the number of stimuli to remember increased to 11 odors, but the inverse relation between accuracy and the number of comparison stimuli was evident throughout the session. Indeed, there was a significant main effect of comparison number [$F(2,10) = 7.72$, $p < .05$], but neither the main effect for bins nor the bin × comparison number interaction was significant ($p > .05$). The middle panel shows accuracy across the 24 sample stimulus sessions; here there was more evidence of a decrease in accuracy as the memory load increased. Although the slope was quite shallow, accuracy dropped from over 90% correct with 0–5 stimuli to remember (bins 1–2) to between 60% and 70% percent correct with 17–23 stimuli to remember (bins 7–8). As in the 12 sample condition, accuracy was also affected by the number of comparisons, with consistently higher levels of performance when there were only two choices. These conclusions are supported by significant main effects for bins [$F(7,35) = 9.57$, $p < .05$], and comparison number [$F(2,10) = 8.06$, $p < .05$], but the bin × comparison number interaction was non-significant. Finally, the bottom panel of Fig. 3 shows within-session data for the 36 sample sessions and again there is some evidence of declining accuracy as the memory load increased. This was particularly evident in the 10-comparison condition for which accuracy declined from over 85% correct in bin 1 to 51% correct in bin 12, but the function was relatively shallow in the 5-comparison condition and quite irregular in the 2-comparison condition. Statistically, however, the outcome was similar to the 24-sample conditions, with significant main effects for bins [$F(11,55) = 3.28$, $p < .05$] and comparison number [$F(2,10) = 26.45$, $p < .05$], and no significant bin × comparison number interaction. Here it should be noted that although percent correct decreased with number of comparisons across the session, the opportunities for error also increased such that in the 5-choice conditions, chance accuracy was .2 from Trial 5 on, and in the 10-choice condition, chance accuracy
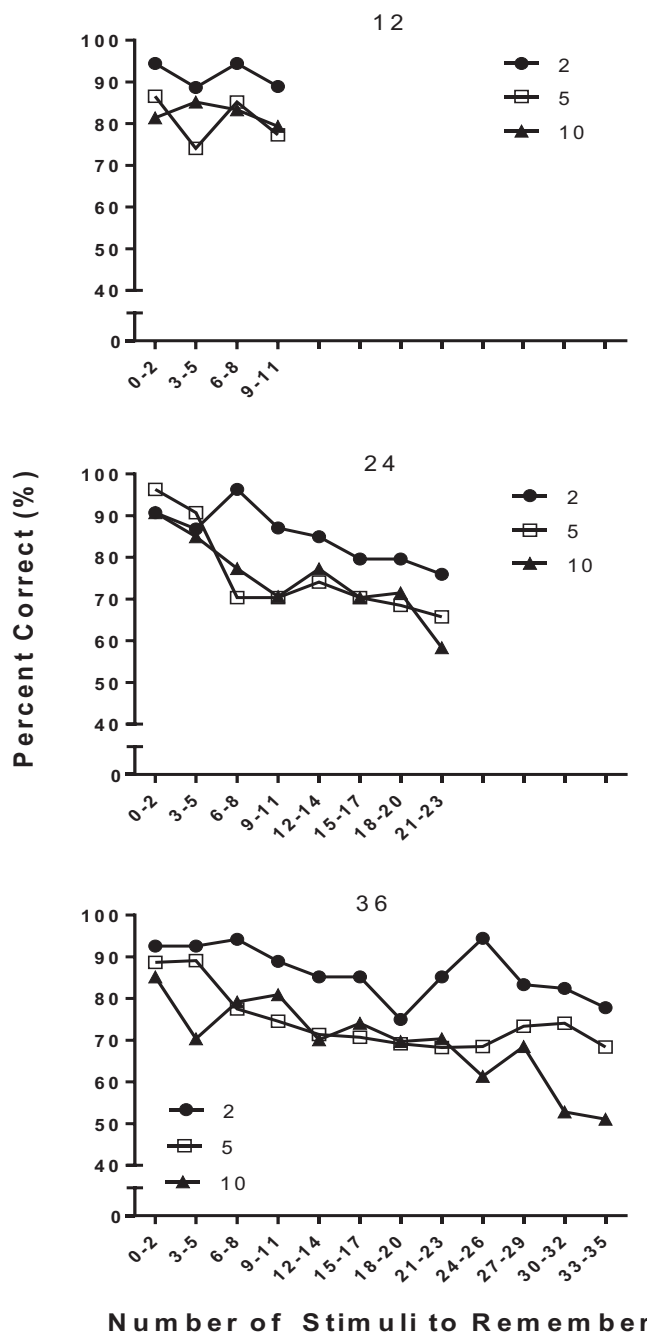
Fig. 3. Percent correct in Experiment 1 as a function of the number of stimuli to remember. Panels depict within session accuracy for conditions involving 12 (top panel), 24 (middle panel), and 36 stimuli (bottom panel). Black circles represent the 2 comparison (comp) choice arrangement, white squares for 5 choices, and black triangles for 10 choices.

was .1 after Trial 10. Thus, although multiple comparisons decreased overall performance, it is clear from Figs. 2 and 3 that accuracies remained at well above chance levels throughout the experiment under all conditions.

Indeed, despite the statistically significant decreases in accuracy observed as the number of stimuli to remember increased, the high overall levels of performance remain striking. For example, under 2-comparison conditions, rats averaged over 80% correct with between 30 and 35 stimuli to remember. Because these high levels of accuracy suggest that even the 36-sample condition did not reach the limits of rats' odor memory capacity, in Experiment 2 we tested animals with longer sessions and up to 72 sample stimuli.

Table 2
Pre-experimental histories for the rats of Experiments 2 & 3.

Rat #	# of sessions	# of stimuli	Stimulus pre-exposure	Stimulus sequence	Experiment	Drug experiments
T12	88	36	No	Ascending	2, 3	Xanomeline
T13	94	36	No	Ascending	2, 3	Xanomeline
S17	172	36	No	Ascending	2	Xanomeline
S1	250	24	No	Ascending	2, 3	MK801, morphine
V20	142	24	No	Ascending	2, 3	Morphine, chlordiazepoxide
D2	137	24	Yes	Random	2	Methamphetamine
E1	163	24	Yes	Random	2	Flunitrazepam, MDMA
E5	106	24	Yes	Random	2	Ketamine, MDMA
F12	110	24	Yes	Random	2	No drugs
F16	113	24	Yes	Random	2	Flunitrazepam
O15	240	36	No	–	3	MK801, morphine, xanomeline
S14	132	36	No	–	3	Xanomeline

Experiment 2: Odor memory capacity in the OST

Experiment 2 was a further attempt to identify capacity limitations in the OST using rats with extensive OST experience in our laboratory. Ten rats tested in previous experiments assessing the effects of drugs on OST performances were studied here under successively longer span tasks (36, 48, and 72 sample stimuli to remember).

Method

Subjects

Ten male HSD rats ≥120 days old at the beginning of OST training served as subjects. All subjects had received extensive prior training in the OST in other studies designed to assess the effects of drugs on OST performance (no Experiment 1 subject was part of Experiment 2). Water and food availability and housing conditions were the same as described for Experiment 1.

Apparatus and stimuli

The arena apparatus was the same as described in Experiment 1 and odor stimuli were prepared in the same way. However, 36 additional odorants were added for a total of 72 (see Table 1).

Procedure

Pre-training for the rats of Experiment 2 was the same as in Experiment 1, but the parameters of OST training always began with 24 stimuli and up to five comparison stimuli (three rats subsequently received sessions with 36 stimuli—see Table 2). Training under these conditions was continued through one or more drug experiments for most rats before Experiment 2 began. Table 2 shows the history of OST training sessions, the number of stimuli used in pre-training, and drug experiments for each subject prior to Experiment 2.

When Experiment 2 began, the number of comparison choices was reduced to two, and the number of stimuli varied between 36, 48 and 72 stimuli. Procedures were the same as those described in Experiment 1 except that the maximum trial duration was 2 min (instead of 1 min). For some subjects (S1, S17, T12, T13, and V20), when tests were conducted with 48 and 72 stimuli, the stimuli added were novel odors. In order to determine whether odor novelty was an important variable, the remaining five rats (D2, E1, D5, F12 and F16) were familiarized with all 72 stimuli before the tests began (see Table 2 for information about stimulus pre-exposure). For these five rats, Experiment 2 began with six OST sessions with 24 stimuli. During this pre-training, rats were exposed to all 72 stimuli balanced for number of presentations across these six sessions. The next session began a series of sessions including 36, 48, or 72 stimuli. All rats were exposed to each stimulus condition one or two times; an ascending sequence was used in some cases and randomized for others.

As in Experiment 1, semi-randomly distributed non-baited control trials were conducted during each session to verify that behavior was under the control of the lid scent and not the sugar pellet. Performance on baited trials was not significantly more accurate than on non-baited trials (mean for baited trials = 88.0; non-baited = 86.2%, $p > .05$). Because of the remarkable performance observed in Experiment 2, an additional control session was conducted after the conclusion of the procedures described above for four of the rats (D2, E1, F12 and F16) to address the possibility that experimenter-cuing or some detectible odor of the baited cups might have influenced responding. This session began as a standard OST procedure, but after 12 trials, control trials were arranged. On these trials, either a reversal was programmed (a previously presented stimulus was reinforced and the "correct" stimulus was not), or neither comparison stimulus was "correct" (both were new or both had been previously presented, but one was arbitrarily selected to be baited). Thus, on these control trials, above chance accuracy would not occur unless behavior was controlled by the presence of the sugar pellet or some other unauthorized cue. Indeed, below chance performances would be expected in the reversal conditions if the lid odors were in full control of responding. The outcomes are shown in Table 3 and the high levels of accuracy on the baseline trials for these sessions are in sharp contrast to the percent selection of the baited stimuli under both control conditions. Under reversal conditions,

Table 3
Percent selection of the baited stimulus on control trials of Experiment 2.

Rat	Baseline	Neither correct	Reversal
D2	100%	38%	13%
E1	100%	38%	13%
F12	92%	56%	25%
F16	83%	44%	13%

rats generally selected the "new" stimulus, in keeping with their training histories, and not the baited stimulus (selected an average of 16% of the reversal trials), whereas mean selection of the baited stimulus occurred on 44% of the trials on which both comparisons were previously presented or both were "new". These data provide additional confirmation that rats' selections were under the control of the lid's odor.

Results and discussion

Fig. 4 shows overall percent correct for individual subjects across the conditions of Experiment 2 and shows remarkable levels of accuracy despite the very high memory loads. Accuracies across sessions ranged from 80.6% to 100% correct, and, although most subjects showed some decrease in accuracy as the number of stimuli increased from 36 to 72, even with 72 stimuli accuracies were consistently well above chance levels. Indeed, overall accuracy ranged from 80.6% to 93.1% on the 72-stimulus sessions and 4 of the 10 subjects obtained higher than 90% correct with 72 stimuli. Mean percent correct was highest on the 36-stimulus sessions ($M = 92.3\%$) and was significantly lower with 48 and 72 stimuli ($M = 88.7$ and 88.4%, respectively), $F(2,16) = 4.65$, $p < .05$. But again, the most striking feature of these data was the high level of accuracy observed even in sessions with up to 72 stimuli to remember.

It should be noted that some of the most accurate performance was observed in the three animals (T12, T13 and S17) who had received pre-training with 36-stimulus (rather than 24-stimulus) OST sessions, particularly at the 36 stimulus condition. This suggests that exposure to these long sessions with more stimuli to remember prior to Experiment 2 may have facilitated subsequent performance with large memory loads. These first three animals, along with S1 and V20, generally showed more accurate performance than the five animals that had pre-exposure (see Table 2) to all 72 stimuli prior to beginning Experiment 2, suggesting that novelty may have increased stimulus saliency. Additionally, Fig. 5 shows within-session performance (blocks of 12 trials) across the three conditions and reveals some decrease in accuracy as a function of the memory load. These decreases across trial blocks were statistically significant for the 36-stimulus [$F(2,19) = 8.52$, $p < .05$] and 72-stimulus [$F(5,54) = 2.70$, $p < .05$] conditions, but not for the 48-stimulus condition ($p > .05$). However, the slopes were very shallow in all three functions, and effects of memory load were clearly less pronounced than those observed in Experiment 1 (cf. Fig. 3).

Span lengths are shown in Table 4 along with the longest run of consecutive correct responses within the session. If the first error in a session (which defines span length) was an index of a memory capacity maximum, the span length should be the same as the longest run, but this was not always the case. Both of these measures were quite variable, with spans ranging from 0 to 47 and longest runs from 9.5 to 47. Longest runs were higher than span length in 19 of the 30 determinations. This finding provides further support for the point noted in Experiment 1 that span length is not the most representative index of OST performance, as rats may make early errors (a span of 0 is caused by an error on Trial 2), yet perform with considerable accuracy throughout much of the remainder of the session. For example, Rat V20 obtained a span of 0 in its 72-stimulus session but had a run of 29 consecutive correct responses later in the session and averaged 91% correct. Similarly, Rat S1 had a span of 1 on its 48-stimulus session, but its longest run was 42 and it completed the session with 95.8% correct. There were no significant differences in either span length or longest run as a function of the number of stimuli to remember ($p > .05$ in both cases).

Table 4
Span length and longest run (LR) for the conditions of Experiment 2.

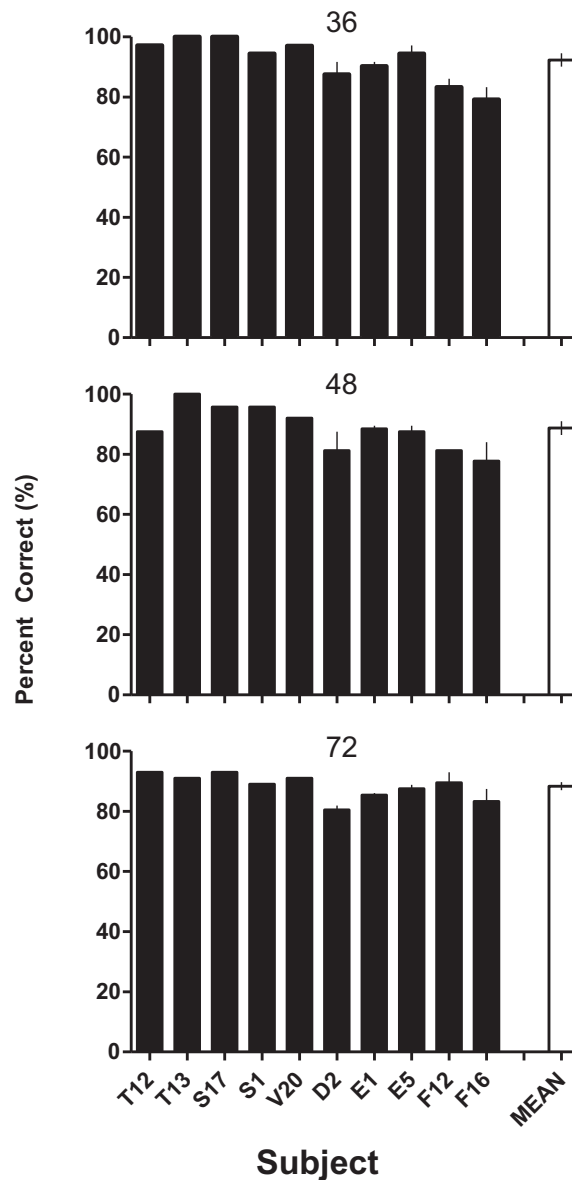| Rat	36 stimuli		48 stimuli		72 stimuli	
	Span	LR	Span	LR	Span	LR
D2	5.0	9.5	6.5	15.5	14.5	14.5
E1	2.5	16.5	13.0	21.0	4.0	23.0
E5	11.0	23.0	10.0	15.0	16.5	16.5
F12	12.5	18.5	8.0	13.0	6.0	23.5
F16	2.5	10.5	2.0	17.0	5.0	15.0
S1	31.0	31.0	1.0	42.0	21.0	21.0
S16	35.0	35.0	28.0	28.0	20.0	20.0
T12	11.0	23.0	14.5	21.0	12.0	28.0
T13	35.0	35.0	47.0	47.0	15.5	18.0
V20	32.0	32.0	20.0	20.0	0.0	29.0
Mean	17.8	23.4	15.0	23.9	11.5	20.9

Fig. 4. Percent correct for individual subjects in Experiment 2, listed in order according to Table 2. Performance on sessions with different numbers of stimuli to remember are shown in each panel: 36 (top panel), 48 (middle panel), and 72 stimuli (bottom panel). Black bars represent individual subject means; white bars depict the grand mean for each condition. For subjects that received two sessions, error bars represent SEM; all other subjects had single sessions at each condition.

The high levels of accuracy obtained in Experiment 2 are all the more impressive when it is noted that for five animals many of the odors presented in the 48- and 72-stimulus sessions were novel. Because accurate performance with novel stimuli in a non-match task like this is often viewed as critical evidence for concept learning (see Bodily, Katz, & Wright, 2008), we also analyzed performance on trials with a novel S+ separately. Overall accuracy on novel S+ trials was 90.3%, and each of the five rats showed accuracies that were significantly above chance (binomial $p < .05$ in each case). These data are thus consistent with previous research in our laboratory showing that rats can learn generalized non-match or oddity concepts with olfactory stimuli (April, Bruce, & Galizio, 2011).

In sum, the effects of memory load demonstrated in previous OST experiments (and Experiment 1 here) were less evident in Experiment 2. Performance was slightly less accurate in 48 and 72 stimulus-span tasks relative to the 36-stimulus task, and there was some decline in accuracy as the number of stimuli to remember increased. However, the striking feature of Experiment 2 was the overall high accuracy across all conditions and the failure to find any indication of a sharp decrease in performance that would suggest a capacity limit—at least up to memory loads of 72 stimuli.
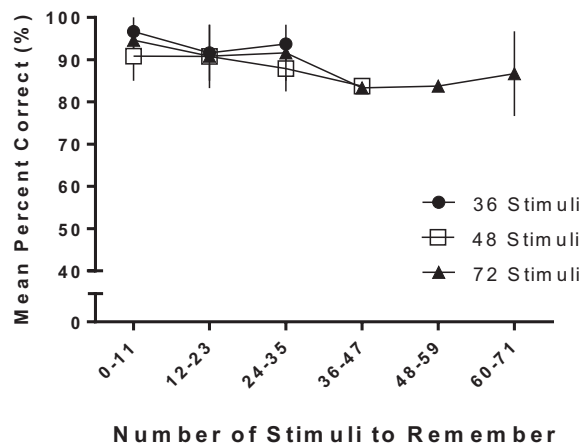
Fig. 5. Mean percent correct in Experiment 2 shown as a function of the number of stimuli to remember (bins of 12 trials). Black circles show performance on 36-stimulus sessions, 48-stimulus conditions are shown as white squares, and 72-stimulus conditions are shown as black triangles. Error bars represent SEM.

Two factors appear critical to the outcomes of Experiment 2. First, the use of only two comparison stimuli may have enhanced performance; recall that in Experiment 1, accuracy was consistently higher and the impact of the memory load was lower with two comparisons. A second factor that appears to be critical is the extended training of the rats in Experiment 2. The number of OST training sessions for rats in Experiment 2 ranged from 88 to 250, and this extensive training is markedly different from most published OST studies and from Experiment 1 in which much less training was administered. This point begs the question of just what types of stimulus control might develop with extended training in the OST that would permit such accurate performance with so many stimuli to remember and over such an extended temporal interval. One possibility is that rats can recognize the relative familiarity of odors in some fashion that is largely independent of the number of stimuli to remember. In other words, within a given session, choices may involve an assessment of the "newness" or "oldness" of a stimulus in relation to the other stimuli present on a given trial, with the ultimate choice being made towards the least familiar option. Thus, with extended training on the OST task, rats might learn to use such familiarity judgments to control stimulus selection. Experiment 3 was designed to explore the sources of control that may be guiding behavior in the OST.

Experiment 3: Relative stimulus familiarity and OST performance

If performance in the OST involves relative judgments of familiarity, then it would seem necessary to sample each comparison stimulus in the array in order to guide response selection. Alternatively, responding could be based on a more absolute identification of odors that had not yet been presented within the session and thus would not require sampling of all stimuli in the array. Such a hypothesis would propose a more detailed type of stimulus control that included both the *specific odor* and *when* it was last presented. There is some evidence of such "what-when" or episodic-like memory for odors in studies using similar tasks (Eichenbaum, Fortin, Ergorul, Wright, & Agster, 2005), and the present study explored these possibilities in two ways.

First, we returned to the data from Experiment 1 and analyzed the video records to determine latency to the first response and number of cups visited. If responding is controlled by relative frequency judgments, we would expect latencies to increase with the number of comparison stimuli in the arena, as each cup should be visited at least once before a response is made.

A second approach to this question was developed by conducting another experiment with a group of six rats with extensive training in the OST (four rats from Experiment 2 and two additional rats). In Experiment 3, we exposed rats to OST sessions with up to five comparison stimuli. On certain designated trials, all five stimuli had previously been presented during the session; there was no true S+ present. These conditions were similar to the control conditions of Experiment 2, but in Experiment 3, five (not two) comparison stimuli were in the arena on probe trials and none of the comparisons were baited. If normal OST performance is guided by selection of an odor not yet presented, then responding should be profoundly disrupted on such trials. On the other hand, if performance is based on relative familiarity, then on these trials rats should select the odor that was presented farthest back in time during the session and with relatively little disruption of responding.

Method

Subjects

Video data from the six rats of Experiment 1 were reanalyzed in the present study. Additionally, the Experiment 3 manipulations were performed on four rats from Experiment 2 and two rats (O15, S14) that had a similar lengthy histories

of OST training, but were not part of Experiment 2 (see Table 2 for details of Rat O15's and Rat S14's pre-experimental training).

Apparatus and stimuli

The arena apparatus was the same as described in Experiments 1 and 2, and odor stimuli were prepared in the same way. Twenty-four odorants from the pool of 72 listed in Table 1 were used.

Procedure

Video records from Experiment 1 were scored for latency to the first response (first lid removed) and number of cups visited. A visit was scored when a rat approached a cup such that its snout was within 1 cm of one of the scented lids without removing the lid.

Experiment 3 began after completion of a baseline session with 24 or 36 stimuli for all six rats (see Table 2). Then sessions were programmed with a mixture of baseline and relative familiarity (RF) probe trials on which all five comparisons had previously been presented during the session. Two different types of relative familiarity probe sessions were conducted.

One type of RF probe session programmed 24 baseline trials using the standard OST procedure (i.e., the number of comparison choices was always five (one S+, four S−)), with six probe trials interspersed during the session set up such that five previously presented odors were randomly selected as comparisons (Random RF probes). None of the cups was baited on these RF probe trials, but the comparison stimulus that had been presented farthest back in time was designated as the "target" stimulus because the relative familiarity hypothesis would predict it to be selected.

The second type of RF probe (Delayed RF probe) session was designed to maximize the time between the first presentation of four designated target odors and their reappearance later in the session. Specifically, the first four stimuli used in the OST were presented alone (i.e., there were no comparison stimuli on these trials) in separate trials and were not used as comparisons until the designated probe trials. After these four stimuli had been presented alone, the OST procedure began. The Delayed RF probe trials began on Trial 13 and occurred every fifth trial thereafter. Each rat was exposed to two sessions of each probe type, and sessions were counterbalanced across rats so that half of the animals received Random RF probe sessions first, whereas the remaining animals received Delayed RF probes first.

Results and discussion

Figs. 6 and 7 show latency to the first response and number of cup visits, respectively, as a function of the number of stimuli to remember for the conditions of Experiment 1. Latency ranged between 10 and 15 s across conditions and was relatively unaffected by either the number of comparison stimuli or the memory load ($p > .05$ in all analyses). In contrast, number of cups visited was clearly a function of the number of comparison cups in the arena with a mean of 0.35 visits per trial in the two-comparison conditions, 1.3 visits in the five-comparison condition, and 2.7 in the 10-comparison condition [$F(2,10) = 726.2$, $p < .05$]. That these striking differences in number of visits occurred without reliable differences in latency indicates speed of responding was determined by multiple factors (e.g., pausing in the arena, variable length of individual visits), in addition to number of visits per se. Although the relative familiarity hypothesis would predict the direction of these results, clearly the mean number of stimuli sampled per trial was considerably fewer than would be expected if an exhaustive search of all of the cups in the arena were required before responding. Additionally, the absence of a latency effect as a function of the number of stimuli to remember is contrary to what would be expected if rats engaged in a serial exhaustive memory search of all stimuli previously encountered in the session (cf., Sternberg, 1966). Rather, these data suggest that rats approached stimulus cups until they encountered the new odor and then generally responded to it, without sampling of all available options. The significant effect of comparison number is consistent with such an explanation because the more stimuli in the arena, the more cups were likely to be sampled before the new odor was encountered. It is also of interest that the number of cups visited was consistently less than would be expected by chance exploration of the arena. Specifically, a random search of all *n* cups until selection of S+ should lead to ($n + 1$)/2 cups searched; however, it is apparent from Fig. 7 that mean visits were consistently below these expected values. This would seem to imply that at least some odors were detected without the close contact that would define a visit in our analysis.

The manipulations of Experiment 3 provided another test of the relative familiarity hypothesis. The top panel of Fig. 8 shows the latency to the first response on RF probe trials of both types and under baseline conditions preceding these probe trials. Latencies on baseline trials were consistently short and, with a mean of 14.1 s, were generally comparable to the baseline latencies observed in Experiment 1 (see Fig. 7). In contrast, both types of RF probes resulted in much longer latencies, with many subjects failing to respond at all on some trials (latencies of 120 s were recorded in such cases). A 2 × 2 factorial ANOVA showed a main effect for trial type (baseline vs. RF probe) [$F(1,5) = 18.54$, $p < .05$], but no differences between session type (Random vs. Delayed RF probe) and no interaction ($F < 1$ in both cases). A similar pattern can be seen in the cup visit data, shown in the middle panel of Fig. 8. Under baseline conditions, visits per trial generally ranged between 1.3 and 2.1, and, as noted above with respect to the Experiment 1 outcomes (Fig. 7), these are fewer than would be expected if rats had to visit each of the five cups before responding on the basis of relative familiarity. In contrast, mean visits per trial on the RF probe trials was 7.9—*more* visits than required to sample each odor. Thus, on trials when all the odors rats encountered had already been presented during the session, they tended to visit stimuli more than once. Again, these conclusions were supported by a main effect of trial type [$F(1,5) = 146.56$, $p < .05$] and no effect of probe type or interaction ($F < 1$ in both cases).
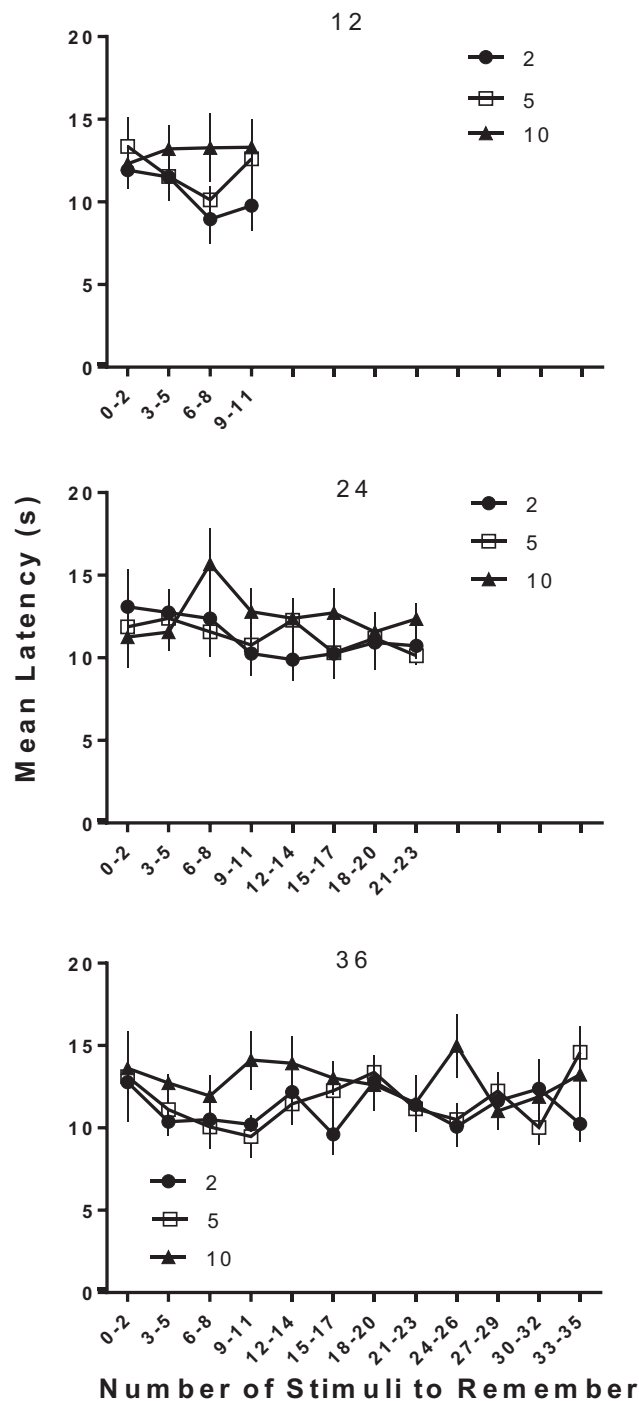
Fig. 6. Mean latencies for 12 (top panel), 24 (middle panel), and 36 (bottom panel) stimulus conditions plotted as a function of the number of stimuli to remember from Experiment 1 (bins of 3 trials). Black circles show performances in the 2-comparison (comp) conditions, white squares for 5 comparisons, and black triangles for 10 comparisons. Error bars represent SEM.

Finally, the bottom panel of Fig. 8 shows the percent of responses made to the target stimulus (the correct/baited S+ on baseline trials and the stimulus presented farthest away in time for RF probe trials). As seen throughout these studies, accuracy was high on baseline trials on both session types, with a mean of over 90% correct. In contrast, selection of the target stimulus occurred at chance levels on Random RF probes ($M = 15.2\%$ target selection, $p > .05$, binomial test). However, target selection rates were somewhat higher and significantly above chance performance on the Delayed RF probe trials ($M = 47.9\%$, $p < .05$, binomial test). On these trials, subjects were more likely to select the temporally most distant of the comparison
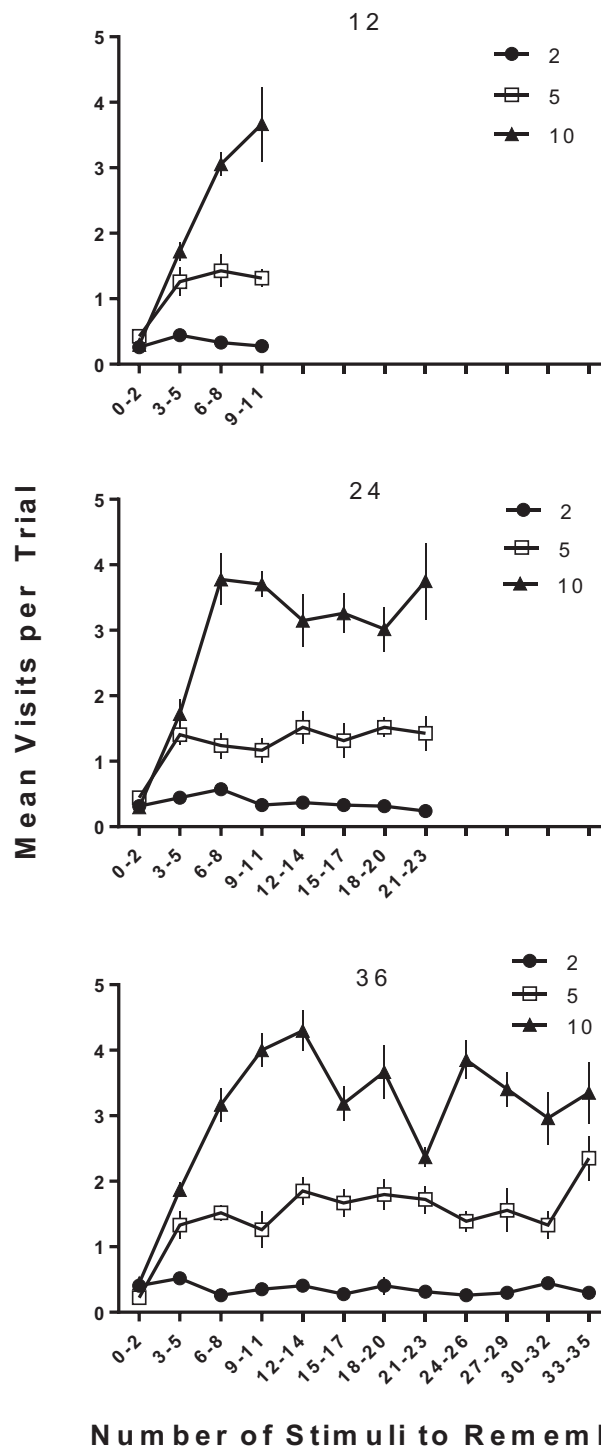
Fig. 7. Mean visits for 12 (top panel), 24 (middle panel), and 36 (bottom panel) stimulus conditions plotted as a function of the number of stimuli to remember from Experiment 1. Black circles show performances in the 2-comparison (comp) conditions, white squares for 5 comparisons, and black triangles for 10 comparisons. Error bars represent SEM.
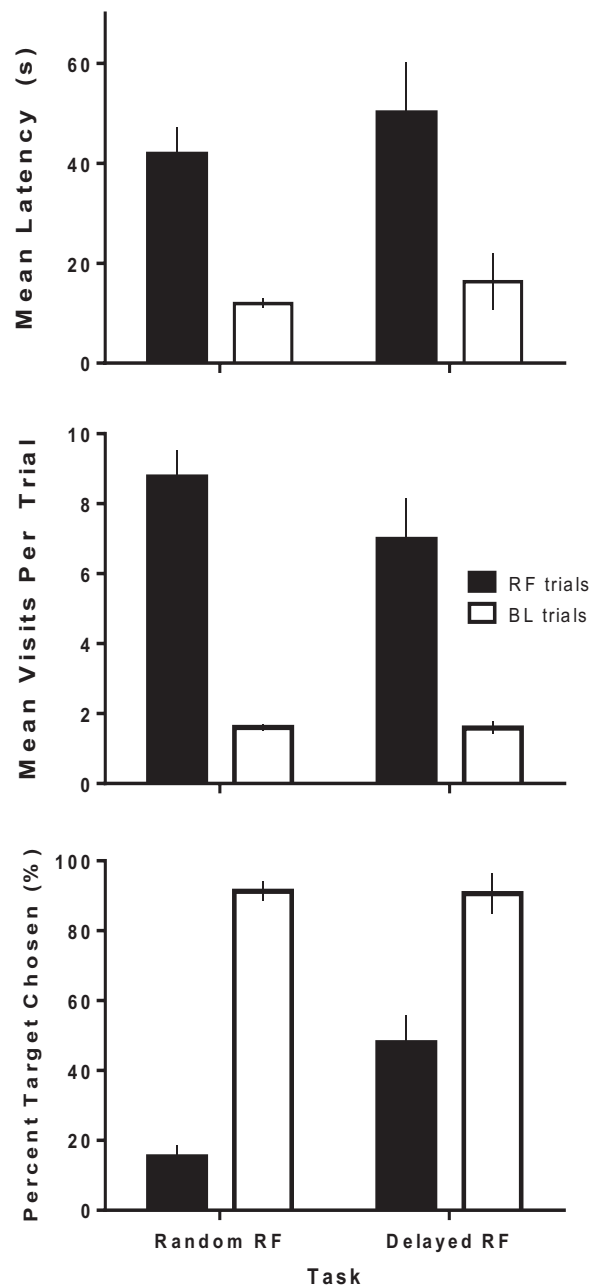
Fig. 8. Experiment 3. Mean latencies (top panel), mean visits per trial (middle panel), and percent target chosen (bottom panel) for both Random and Delayed Relative Familiarity Probe tasks. Performances on Relative Familiarity probe trials (RF) are shown with black bars while Baseline trials (BL) are white bars. Error bars represent SEM.

stimuli, although not at rates that approached accuracy under baseline conditions. A significant session type × probe type interaction [$F(1,5) = 20.13$, $p < .05$] confirmed these conclusions. This interaction is based on the higher rates of target selection on the Delayed RF probe trials and provides some evidence that responding in the OST can be controlled, at least in part, by relative familiarity.

General discussion

The OST has generally been viewed as an assessment of rats' working memory capacity, but the present results raise questions about this interpretation. In Experiment 1, we found that the number of comparison stimuli clearly influenced accuracy, a variable which is confounded with number of odors to remember in most previous OST experiments (Dudchenko

et al., 2000; Rushforth et al., 2010, 2011; Turchi & Sarter, 2000; Young, Crawford et al., 2007; Young, Kerr et al., 2007; Young et al., 2008). Experiment 1 also showed an independent effect of the number of sample stimuli to remember, as would be expected in an index of memory capacity. However, accuracies remained well above chance in Experiment 1 even with 36 stimuli to remember; this finding led to the analysis of animals with extensive OST experience under even higher memory loads in Experiment 2. That these animals performed at high levels of accuracy with as many as 72 stimuli to remember was quite remarkable as a demonstration of the extent to which multiple stimuli can maintain control of behavior in the rat within a single session. Various control conditions confirmed that it was indeed the lid odors that were controlling responding and not the scent of the sucrose reinforcer or other unauthorized variables.

However, it is clear from the high levels of accuracy observed in Experiment 2 that the capacity for remembering odors in well-trained rats is not limited to 72 stimuli. There may well be an upper limit to the number of odors rats can remember in the OST, but if so, it was not reached in the present study. We did not increase the number of stimuli beyond 72 in the present study because of the length and complexity of the session that would be required, but such an experiment seems worth exploring despite the challenges. It is possible though that odor recognition memory for rats may be much like recognition memory for meaningful pictures in humans, with such a large capacity that it appears unlimited (cf. Standing, 1973).

Indeed, the very high capacities displayed in the present study suggest that the OST is measuring a type of remembering that is quite different from that of the classic working memory tasks used with human participants (e.g., digit span and n-back, Gathercole, 2009). Previous interpretations of very limited memory capacities in studies using the OST may have been augmented by the use of span length as the primary dependent variable (e.g., Rushforth et al., 2010, 2011). However, we found span length to under-represent performance in Experiments 1 and 2 in several ways. Span lengths varied widely, but within-session analysis revealed accuracies that were well above chance even late in the session—well beyond the span length. Also, the longest run in the session was often not the initial span length. That is, making an initial error was not indicative of a capacity limitation in that subsequent performance continued to show high levels of accuracy. This is quite unlike performance in tasks such as the digit span, which typically show complete failure once the span length is reached (Gathercole, 2009). Thus, studies that have interpreted variables altering span as affecting working memory capacity may bear reconsideration. Even studies that have examined overall OST accuracy as well as span length may not be simply interpreted in terms of working memory capacity (e.g., Cui et al., 2011; Dudchenko et al., 2000; Young, Crawford et al., 2007; Young, Kerr et al., 2007; Young et al., 2008) because of the confound between the number of stimuli to remember and number of comparison stimuli in the test arena. The effects of the number of comparison stimuli observed in the present study indicate that the number of comparison stimuli must be held constant through the session in order to permit interpretation in terms of capacity effects (Galizio et al., 2013; MacQueen et al., 2011).

The high levels of accuracy observed in Experiments 1 and 2 led to questions about the sources of stimulus control in the OST and particularly whether performance might be controlled by relative odor familiarity. Rats have an innate preference for novel or unfamiliar objects [which is the basis for the novel object task (Ennaceur & Delacour, 1988)], but this preference itself is not sufficient to account for OST responding as several sessions of training were required to produce accurate performances on the OST. Still, Experiment 3 provided some support for the idea that relative familiarity may be one factor guiding OST responding. In particular, the difference between rates of target selection on the random vs. delayed probes seems to be explicable in terms of relative familiarity. Target comparisons in the random condition were often just one trial removed from some of the other comparisons in the array on a given trial, but the target stimulus on a delayed probe was removed by 7–23 trials from any of the other comparisons present on an RF probe trial. This difference would seem to be the basis for the difference in target selection between the two probe types and supports the idea that stimulus control by relative familiarity can play a role in OST performances.

However, some aspects of the data do not seem amenable to a relative familiarity interpretation. For example, target selection of delayed probes ranged between 25% and 63% and was thus well below accuracy obtained during baseline OST trials on which one stimulus had not yet been presented during the session. Perhaps this difference could still be accounted for in terms of relative familiarity if it is argued that baseline trials represent a very large difference in relative familiarity (24 hr vs. a few min). Thus, perhaps the striking differences in relative familiarity on baseline trials simply present an easier discrimination with correspondingly greater accuracy. However, latency and number of visits further suggest that more than relative familiarity may be controlling OST responding. There was no difference in latency or number of visits between the two probe types, but both RF probe trials showed much longer latencies and more cup visits than was seen on baseline trials. Clearly rats were behaving quite differently on these RF probe trials on which all comparison stimuli already had been presented during the current session. The basis for this difference seems to be that on baseline trials rats appear to recognize the new/correct stimulus when they encounter it and respond without visiting additional stimuli. In contrast, on RF probe trials stimuli were generally visited more than once and rats sometimes failed to respond to any stimulus even after sampling them all. So it appears that relative familiarity may affect responding on probe trials when no new stimulus is available, but that on typical baseline trials responding is determined more specifically by whether a particular odor had previously been encountered during the current session or whether it was novel to the session. Both lid odor and the time the odor was last encountered appear to jointly control OST responding and this "what-when" stimulus control may represent a form of episodic-like remembering. Such a conclusion is consistent with observations using similar odor memory procedures and signal detection methodologies by Eichenbaum and colleagues that have successfully dissociated familiarity and recollection processes (Eichenbaum et al., 2005; Sauvage, Beer, & Eichenbaum, 2010), but further analysis is needed to explore this possibility in the OST. For example, in the present study it is not possible to determine whether

selection by the rats was based on remembering specifically *when* they encountered a particular odor or whether it was based on simply remembering that a long time has passed since encountering that particular stimulus. Methodologies for testing these hypotheses have been developed and could be applied to the OST (Roberts et al., 2008; Zhou & Crystal, 2011).

Finally, it should be noted that although the present data suggest that the OST does not appear to assess a highly limited odor memory capacity in the rat, it still meets the generally stated criteria that are used to define working memory tasks in non-humans. That is, in the OST rats remember which odors they have been exposed to *within* each experimental session, but not *between* sessions (Dudchenko, 2004). This oft-used definition results in the classification of tasks such as the Radial Arm Maze, some versions of the Morris Swim Task, the Object Recognition task, and Delayed Match to Sample as working memory tasks. Of these tasks, only the OST was specifically designed to study capacity limits and has typically been interpreted in these terms. However, given the present failure to find such limits with a task having the high face validity of the OST, perhaps the translational value of the widely accepted definition of non-human working memory may be called into question more generally (note also Cole & Chappell-Stephenson, 2003; Roberts, 1979). If working memory tasks in non-humans are assessing a different set of memory processes than those we refer to as working memory in humans, it seems important to further identify them. This point is of more than passing interest because neurobiological variables affecting performances on these tasks are coming to be increasingly well understood (cf., Dudchenko, Talpos, Young, & Baxter, 2012), and it is of considerable importance to be able to relate these findings to human memory processes.

Acknowledgements

References

Alvarado, M. C., & Bachevalier, J. (2008). Animal models of amnesia. In H. Eichenbaum (Ed.), *Learning and memory: A comprehensive reference* (pp. 143–168). Amsterdam: Elsevier Press.

April, L. B., Bruce, K., & Galizio, M. (2011). Matching- and non-matching-to-sample concept learning in rats using olfactory stimuli. *Journal of the Experimental Analysis of Behavior*, 96, 123–138.

Bodily, K. D., Katz, J. S., & Wright, A. A. (2008). Matching-to-sample abstract-concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 178–184.

Cole, M., & Chappell-Stephenson, R. R. (2003). Exploring the limits of spatial memory in rats using very large mazes. *Learning & Behavior*, 31, 349–368.

Cui, Y., Jin, J., Zhang, X., Xu, H., Yang, L., Du, D., et al. (2011). Forebrain NR2B overexpression facilitating the prefrontal cortex long-term potentiation and enhancing working memory function in mice. *PLoS ONE*, 6, 1–10.

Dudchenko, P. A. (2004). An overview of the tasks used to test working memory in rodents. *Neuroscience & Biobehavioral Reviews*, 28, 699–709.

Dudchenko, P. A., Talpos, J., Young, J., & Baxter, M. G. (2012). Animal models of working memory: A review of tasks that might be used in screening drug treatments for the memory impairments found in schizophrenia. *Neuroscience & Biobehavioral Reviews*, http://dx.doi.org/10.1016/jneubiorev.2012.03.003

Dudchenko, P. A., Wood, E. R., & Eichenbaum, H. (2000). Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. *Journal of Neuroscience*, 20, 2964–2977.

Eichenbaum, H., Fortin, N. J., Ergorul, C., Wright, S. P., & Agster, K. L. (2005). Episodic recollection in animals: "If it walks like a duck and quacks like a duck. . .". *Learning and Motivation*, 36, 190–207.

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. I. *Behavioural Brain Research*, 33, 197–207.

Galizio, M., Deal, M., Hawkey, A., & April, L. B. (2013). Working memory in the odor span task: Effects of chlordiazepoxide, dizocilpine (MK801), morphine, and scopolamine. *Psychopharmacology*, 225, 397–406.

Gathercole, S. E. (2009). Working memory. In J. H. Byrne (Ed.), *Concise learning and memory: Editor's selection* (pp. 149–168). Amsterdam: Elsevier Press.

MacQueen, D. A., Bullard, L., & Galizio, M. (2011). Effects of dizocilpine (MK801) on olfactory span in rats. *Neurobiology of Learning and Memory*, 95, 57–63.

Roberts, W. A. (1979). Spatial memory in the rat on a hierarchical maze. *Learning and Motivation*, 10, 117–140.

Roberts, W. A., Feeney, M. C., MacPherson, K., Petter, M., McMillan, N., & Musolino, E. (2008). Episodic-like memory in rats: Is it based on when or how ago? *Science*, 320, 113–115.

Rushforth, S. L., Allison, C., Wonnacott, S., & Shoaib, M. (2010). Subtype-selective nicotinic agonists enhance olfactory working memory in normal rats: A novel use of the odour span task. *Neuroscience Letters*, 471, 114–118.

Rushforth, S. L., Steckler, T., & Shoaib, M. (2011). Nicotine improves working memory span capacity in rats following sub-chronic ketamine exposure. *Neuropsychopharmcology*, 36, 2774–2781.

Sauvage, M. M., Beer, Z., & Eichenbaum, H. (2010). Recognition memory: Adding a response deadline eliminates recollection but spares familiarity. *Learning & Memory*, 17, 104–108.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207–222.

Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652–654.

Turchi, J., & Sarter, M. (2000). Cortical cholinergic inputs mediate processing capacity: Effects of 192 IgG-saporin-induced lesions on olfactory span performance. *European Journal of Neuroscience*, 12, 4505–4514.

Wright, A. A. (2007). An experimental analysis of memory processing. *Journal of the Experimental Analysis of Behavior*, 88, 405–433.

Young, J. W., Crawford, N., Kelly, J. S., Kerr, L. E., Marston, H. M., Spratt, C., et al. (2007). Impaired attention is central to the cognitive deficits observed in alpha 7 deficient mice. *European Neuropharmacology*, 17, 145–155.

Young, J. W., Kerr, L. E., Kelly, J. S., Marston, H. M., Spratt, C., Finlayson, K., et al. (2007). The odour span task: A novel paradigm for assessing working memory in mice. *Neuropharmacology*, 52, 634–645.

Young, J. W., Sharkey, J., & Finlayson, K. (2008). Progressive impairment in olfactory working memory in a mouse model of Mild Cognitive Impairment. *Neurobiology of Aging*, 30, 1430–1443.

Zhou, W., & Crystal, J. D. (2011). Validation of a rodent model of episodic memory. *Animal Cognition*, 14, 325–340.