

Humans and monkeys use different strategies to solve the same short-term memory tasks

John H. Wittig Jr.,¹ Barak Morgan,^{2,3} Evan Masseau,¹ and Barry J. Richmond¹

¹Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892-4415, USA; ²Global Risk Governance Program, Department of Public Law, University of Cape Town, Rondebosch 7701, South Africa; ³DST-NRF Centre of Excellence in Human Development, DVC Research Office, University of Witwatersrand, Johannesburg, 2050, South Africa

The neural mechanisms underlying human working memory are often inferred from studies using old-world monkeys. Humans use working memory to selectively memorize important information. We recently reported that monkeys do not seem to use selective memorization under experimental conditions that are common in monkey research, but less common in human research. Here we compare the performance of humans and monkeys under the same experimental conditions. Humans selectively remember important images whereas monkeys largely rely on recency information from non-selective memorization. Working memory studies in old-world monkeys must be interpreted cautiously when making inferences about the mechanisms underlying human working memory.

[Supplemental material is available for this article.]

Humans use working memory to selectively store important information while ignoring or forgetting distracting information. Human working memory is an attention-demanding storage mechanism for mental rehearsal and/or manipulation of information (Cowan 2008; Baddeley 2010). Working memory is commonly studied in animals by testing whether they recognize recently viewed images. In such tasks, an image that should be remembered is a “target” and an image that should be forgotten or ignored is a “distractor.” One or more images are presented in sequence followed by a test image that matches a target, matches a distractor, or does not match any image from the sequence. Successful discrimination between a test image that matches a target versus a distractor is often assumed to depend on mental rehearsal of the target using working memory (Miller and Desimone 1994; Scott et al. 2012).

We recently reported that seven monkeys performed well on several working memory tests using a strategy that circumvented the need for selective memorization of target images (Wittig and Richmond 2014). Our conclusion was based on an analysis of response rates (hits and false alarms) as a function of how recently a test image was last presented. In line with previous reports, hit rates increased when a test image matched a recently presented image, the so-called recency effect (Yakovlev et al. 2005; Wright 2007). To our surprise, false alarm rates also appeared to be governed by the same recency effect, such that response rates depended on how recently a test image was last presented, not whether the test image matched a target or distractor. Thus, it appeared as though our monkeys indiscriminately memorized targets and distractors, and then performed the task with high overall accuracy simply by judging how recently each test image was last seen, or how familiar it was.

We argued that our monkeys' use of recency information, applied indiscriminately to targets and distractors, could be supported by a memory mechanism akin to human familiarity. Familiarity supports recognition of a large number of items, is ef-

fortless to use, and conveys a unitary “strength” signal that follows the laws of recency and repetition, where items seen more recently or more frequently have a stronger memory signal (Eichenbaum 2008; Kahana 2012). We conjectured that monkeys prefer using an effortless memory mechanism such as familiarity instead of an attention-demanding memory mechanism such as working memory. Our previous finding in monkeys left an open question: would humans also use a nonselective memory mechanism if experimental conditions matched those in monkey experiments?

Here we compare new data from 20 humans to a mix of new and previously reported data from five rhesus monkeys as they performed three recognition memory tasks (Fig. 1). For each task, a sequence of images was presented on a computer monitor, one at a time, and the sequence ended with a “yes/no” recognition test. A subject had to report whether or not the test image matched a target image from that sequence. For each sequence, there was a 50% chance that the test matched a target. A common manipulation in recognition memory tasks is to reuse a small set of images across trials so that a target from one trial can be a distractor in the next, thus causing high interference and presumably limiting the use of familiarity to solve the task (Basile and Hampton 2013). For each task variant, we used both a large (100 images) and small (two or five images) stimulus set to see whether this manipulation differentially affected humans and monkeys.

In the first and simplest task, Same–Different, there were two images per sequence—the target and the test. When the test matched the target the correct response was “yes” (Fig. 1A, *Trial #1*). When the test did not match the target the correct response was “no” even though the test might have matched an image from a previous sequence, which we call a *preceding distractor* (Fig. 1A, *Trial #2* and *#3*). Both humans and monkeys perform

© 2016 Wittig, et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see <http://learnmem.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding author: bjr@ln.nimh.nih.gov

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.041764.116>.

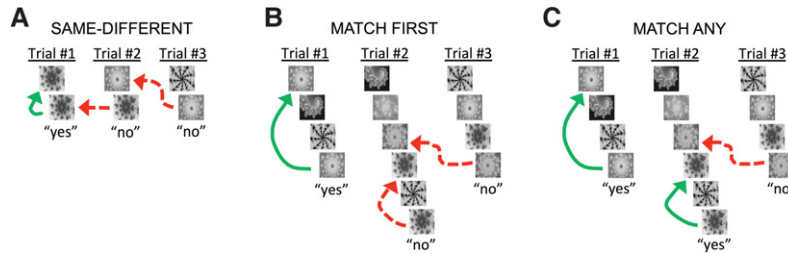


Figure 1. Example trials from the three recognition memory tasks: Same–Different, Match First, and Match Any. Sequence length was 2 for Same–Different, and 4 or 8 for Match First and Match Any (6 shown for illustration purposes). The last image of a sequence is the test image. A green square appeared in the center of the screen (not shown in illustration) to cue subjects to respond “yes” by releasing a touch bar (for monkeys) or space bar (for humans) within 1 sec, or “no” by continuing to hold until the green square disappears. A correct response to the test image (shown below each trial) depends on whether the test matches a target image (“yes” correct) or a distractor image (“no” correct), as defined by the task rule. Solid green arrows indicate which target was tested and dashed red arrows indicate which distractor was tested. (A) In Same–Different, a yes response is correct if and only if the test matches the only other image in the sequence. (B) In Match First, a yes response is correct if and only if the test matches the first image from the current sequence. (C) In Match Any, a yes response is correct if and only if the test matches any image from the current sequence. Each task was inspired by previously described memory tests for monkeys, where Same–Different mimics “delayed-match-to-sample” (Bauer and Fuster 1976; Mishkin and Manning 1978), Match First mimics “ABBA” (Miller and Desimone 1994; Miller et al. 1996), and Match Any mimics Serial Probe Recognition (Wright 2007; Basile and Hampton 2010). Monkeys were trained and tested over the course of months, using color images of natural scenes and clip art cartoons, with images and delays between 500 and 2000 msec depending on task condition (Same–Different and Match Any images and delays were 1000 msec; Match First target images were 2000 msec, distractors were 500 msec, and delays were 750 msec). Humans were trained and tested in a single 90-min session, using gray-scale fractal images to minimize the use of verbal memory (shown with contrast increased 40% to improve print quality), with 500 msec image and delay durations. Over the course of a 90-min session, the software automatically progressed through training and testing phases for each task in the same order used with the monkeys. Details of monkey training and testing have been reported previously (Wittig and Richmond 2014) and are described in comparison to the human training and testing protocol in Supplemental Experimental Methods.

well on Same–Different, with humans performing better (Fig. 2A). Reducing the size of the stimulus set from 100 to two images caused a comparable drop in performance for monkeys and humans (Fig. 2A, no significant species × set size interaction).

In the second task, Match First, each sequence had four or eight images. The first image in each sequence was the target and the remaining images before the test were *intervening distractors*. When the test matched the target, the correct response was “yes” (Fig. 1B, Trial #1). When the test did not match the target the correct response was “no,” even though the test might have matched an *intervening distractor* from the current sequence (Fig. 1B, Trial #2), or a *preceding distractor* from a previous sequence (Fig. 1B, Trial #3). Humans performed well on Match First for all set sizes and sequence lengths. Monkeys performed poorly, with performance becoming worse as set size decreased or the number of interfering distractors increased (Fig. 2B, significant species × set size, and species × distractor interactions).

In the third task, Match Any, again there were 4 or 8 images per sequence, but in this case every image before the test was a target. When the test matched any of the targets the correct response was “yes” (Fig. 1C, Trials #1 and #2). When the test did not match any of the targets the correct response was “no,” even though the test might have matched a *preceding distractor* from a previous sequence (Fig. 1C, Trial #3). Human and monkey performance was comparable on Match Any for all set sizes and sequence lengths (Fig. 2C, no main effect of species, no significant interactions).

If we had only seen performance from Same–Different and Match Any, we might have concluded that monkeys and humans use similar strategies. However, the large differences in performance in Match First warranted a closer examination of how each task was solved. For each task variant, we produced a serial position curve by plotting “yes” response rates as a function of how recently a test image was previously presented. We adopt the terminology of signal detection theory (Wickens 2002), where hit rate is the percentage of *correct* “yes” responses when the test matches a target (hits/(hits + misses)), and false alarm rate is the percentage of *incorrect* “yes” responses when the test matches a distractor (false alarms/(false alarms + correct rejections)).

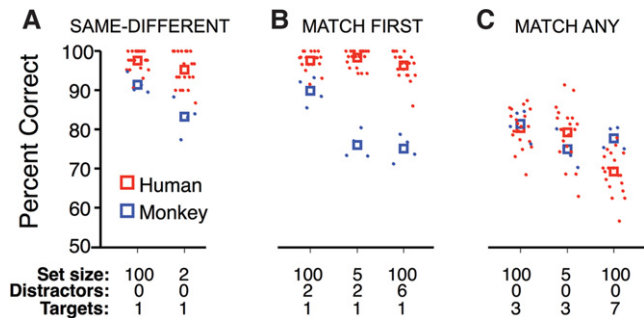


Figure 2. Humans and monkeys are sensitive to different task manipulations. Performance (% correct over all trials) of individual human and monkey subjects for each task condition are shown as red and blue points, respectively, with group means indicated with open squares. Performance is presented separately for each task in A–C, with specific conditions of set size, number of intervening distractors per sequence, and number of targets per sequence enumerated below the abscissa. We examine how task manipulations differentially affected humans and monkeys by evaluating five separate repeated-measures, two-way ANOVA models, each with species as one factor. Separate models were required due to incomplete crossing of experimental factors in the data set. We tested for the main effect of set size in each task (three models), number of intervening distractors in Match First (one model), and number of targets in Match Any (one model). Significance of each test was corrected for multiple comparisons (Bonferroni correction, $P = 0.05/5 = 0.01$). Percent correct data were normalized using the modified arcsine transform (Freeman and Tukey 1950; Zar 2010). (A) In Same–Different, there was a main effect of species ($F_{(1,22)} = 17.8, P = 0.0004$) and set size ($F_{(1,22)} = 8.3, P = 0.009$), but no interaction ($F_{(1,22)} = 1.0, P = 0.324$). (B) In Match First, there was a main effect of species ($F_{(1,22)} = 118.9, P = 2.5e-10$), no main effect of set size ($F_{(1,22)} = 0.06, P = 0.81$), and a significant interaction ($F_{(1,22)} = 13.0, P = 0.002$). There was also a main effect of the number of intervening distractors per sequence ($F_{(1,22)} = 9.7, P = 0.005$) and an interaction between species and number of intervening distractors ($F_{(1,22)} = 11.8, P = 0.002$). (C) In Match Any, there was no main effect of species ($F_{(1,22)} = 0.2, P = 0.645$), no main effect of set size ($F_{(1,22)} = 1.9, P = 0.188$), and no interaction ($F_{(1,22)} = 2.6, P = 0.118$). There was a main effect of number of targets per sequence ($F_{(1,22)} = 95.8, P = 1.8e-09$), but no interaction ($F_{(1,22)} = 6.6, P = 0.018$).

In Same-Different, the hit rate was high for both monkeys and humans (Fig. 3A,B, *red and blue squares at Test Recency = -1*). However, the false alarm rates were strikingly different for the two species. Humans made almost no false alarms (Fig. 3A,B, *red squares at Test Recency = -5 to -2*), whereas monkeys made an increasing proportion of false alarms for more recent distractors (Fig. 3A,B, *blue squares at Test Recency = -3 and -2*). For monkeys performing Same-Different with a large stimulus set, a single straight line described the relationship between “yes” response rate and test recency irrespective of whether the test matched a target or distractor (Fig. 3A, *blue line from Test Recency = -3 to -1*). Despite the monkeys’ tendency to make false alarms for recently presented distractors, their overall performance was high, over 90% (Fig. 2A). For monkeys performing Same-Different with a small stimulus set (Fig. 3B, *blue squares*), hit and false alarm rates increased with test recency, but unlike in Figure 3A, hit rate (Test Recency = -1) did not fall along the line connecting false alarm rates (Test Recency = -3 and -2).

The serial position curves of humans and monkeys performing Same-Different illustrate the signatures of recency-based and target-selective behavioral strategies. The signature of the

recency-based strategy is that response rates depend on how recently a test image was last presented, irrespective of whether the test matched a target or distractor; hit and false alarm rates fall along the same line (Fig. 3A, *blue squares and line*). In contrast, the signature of the target-selective strategy is that response rates depend exclusively on whether the test matches a target versus a distractor, irrespective of test recency; hit and false alarm rates do not fall along the same line and false alarm rates do not change with test recency (Fig. 3A,B, *red squares and lines*). A combination of these strategies is evident when false alarm rates change systematically with test recency, but hit rates and false alarm rates do not fall along the same line (Fig. 3B, *blue squares and line*). Based on these criteria, humans exclusively use a target-selective strategy when performing Same-Different, whereas monkeys use either a recency-based strategy (for a large stimulus set) or a mix of recency-based and target-selective strategies (for a small stimulus set).

In Match First, monkeys had high false alarm rates for distractors presented just before or after the target stimulus (Fig. 3C,D, *blue squares at Test Recency = -4 and -2*; Fig. 3E, *blue squares at Test Recency = -8 and -6*). With two intervening distractors and a small stimulus set, a single straight line described the relationship between “yes” response rates and

test recency irrespective of whether the test matched a target or distractor (Fig. 3D, *blue line from Test Recency = -3 to -1*). In contrast, humans just about never made false alarms in Match First (Fig. 3C-E, *red squares*). Human false alarm rates did show a slight but significant dependence on test recency with six intervening distractors (Fig. 3E, *red line from Test Recency = -6 to -1*). Nonetheless, the false alarm rate for the human subjects was extremely low. In Match First, monkeys relied heavily, sometimes exclusively, on a recency-based strategy. Humans relied nearly exclusively on a target-selective strategy.

In Match Any, monkey response rates depended linearly on test recency, irrespective of whether the test matched a target or distractor (Fig. 3F,G, *blue lines from Test Recency = -5 to -1*; Fig. 3H, *blue line from Test Recency = -9 to -1*). Humans treated each of the three conditions of Match any differently. With three targets and a large stimulus set, human response rates depended on test recency for targets, but not distractors (Fig. 3F; *red line from Test Recency = -5 to -4*). With three targets and a small stimulus set, human response rates depended on test recency for both targets and distractors, but a single straight line did not fit all of the data; we used different lines for targets and distractors (Fig. 3G, *red line from Test Recency = -5 to -4 does not predict responses at Test Recency = -3*). Finally, for Match Any with seven targets and a large stimulus set, human response rates depended linearly on test recency using a single line for hits and false alarms (Fig. 3H, *red line from Test Recency = -9 to -1*). In Match Any, monkeys rely exclusively on a recency-based strategy,

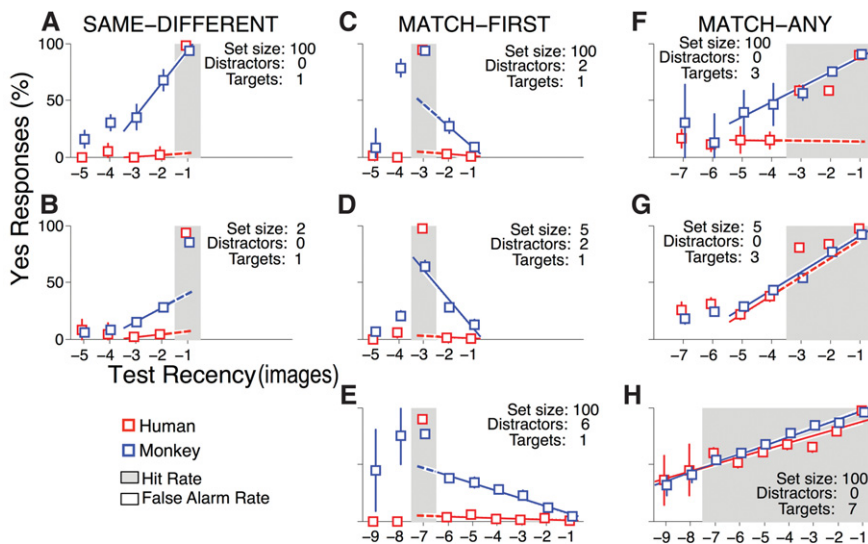


Figure 3. Testing for recency-based versus target-selective behavioral strategies. For each task condition, a serial position curve was created by sorting trials according to how recently a test image was previously presented, in terms of number of images prior, and then calculating yes response rates for each test recency. The average response rates for humans and monkeys are indicated by red and blue squares, respectively, and error bars are \pm SEM (20 humans, 4 monkeys). Gray backgrounds indicate test recency for trials where the test matched a target (yes correct, hit rates reported) and white backgrounds indicate test recency for trials where the test matched a distractor (yes incorrect, false alarm rates reported). For instance, in Same-Different (A,B), a yes response is correct only if the test matches the immediately preceding image at Test Recency = -1, whereas in Match First, a yes response is correct only if the test matches the first image in the sequence, which is at Test Recency = -3 when there are two intervening distractors (C,D) or Test Recency = -7 when there are six intervening distractors (E). Red and blue lines in each panel are linear regressions on the human and monkey data. For each task condition and species, we first attempted to fit a single linear regression to hit rates and false alarm rates. Based on an ANOVA of the residuals (Zar 2010), a single linear regression was appropriate for monkeys in A,D,F-H, and for humans in H (solid lines indicate range of the regression). If a single linear regression failed to account for hit rates and false alarm rates, we next regressed on false alarm rates for the two most recent preceding distractors (for Same-Different and Match Any) or all intervening distractors (for Match First) and projected these regressions to the test recency where a yes was correct. For instance, in Same-Different (B) the regression on distractors at Test Recency = -3 to -2 (solid lines) is projected to the targets at -1 (dashed lines). Thus, solid lines connecting targets and distractors indicate collinearity of false alarms and hits and represent a pure recency strategy. In contrast, dashed lines indicate that hits are not collinear with false alarms, and if the slope is zero, represent a pure target-selective strategy. Dashed lines that have a nonzero slope represent a combination of recency-based and target-selective strategies. For every task condition (A-H), the regression on the monkey data had a nonzero slope (ANOVA $P < 0.05$; Zar 2010). For humans, only the regressions in E,G,H had nonzero slopes.

whereas humans use a mix of target-selective (Fig. 3F,G) and recency-based (Fig. 3G,H) strategies.

We asked our human subjects to report on a scale of 1–10 how often they had been “repeating the name [they gave the image] or mentally picturing the image” in each task, where 1 was “never rehearsed” and 10 was “rehearsed every image.” Most subjects mentally rehearsed the images in Match Any (17 of 20 responses >1; mean response 6 ± 1) and all subjects rehearsed the first image in Match First (20 of 20 responses >1; mean response of 8 ± 1 is significantly greater than that for Match Any, paired *t*-test, $t_{19} = -2.3$, $P = 0.033$). The working memory system supports conscious rehearsal of target stimuli (Logie 2011; Baddeley 2012). Thus, these subjective reports are consistent with the notion that these tasks are solved using working memory to selectively remember the targets, at least in humans.

Other authors have recognized that typical “working memory” tests may not measure the same processes in monkeys and humans, despite the tests being at least superficially similar across species (Cook et al. 1991; Basile and Hampton 2013). Our data illustrate the task conditions under which monkey memory processes do not “map” well to human working memory. Mapping failed when humans used a target-selective strategy but monkeys did not. Humans used a target-selective strategy in all but one condition (seven of eight conditions). Mapping failed in four of those seven conditions when monkeys used a pure recency-based strategy. Even when monkeys did use a target-selective strategy, it was always in combination with a recency-based strategy. Thus when mapping did not fail, it was imperfect. In general, monkey memory processes did not map well to human working memory.

Humans were much better at selectively remembering targets, which could reflect the use of language to recode the images with names that are easy to rehearse, a superior understanding of the task rules, or superior visual working memory capacity and selectivity. Humans accurately and selectively remembered target images even though it would be difficult to apply a unique name to 100 different gray-scale fractal images. The only case of humans not using a target-selective strategy occurred during Match Any with seven target images, a condition where the number of memorized images might exceed the capacity of human working memory. Human subjects clearly understood the task rule for Match Any, based on their performance with three target images, yet they resorted to a nonselective strategy with seven target images. So for humans, understanding the task rule did not guarantee use of a target-selective strategy. We were initially concerned that monkeys’ lack of a target-selective strategy was due to confusion about the different task rules because of our training regiment, but we confirmed that the recency-based strategy still dominated when training order was changed (see Supplemental Materials and Methods). This suggests that the primary difference between species is not related to the use of verbal coding, nor understanding the task rules, but rather reflects a difference in the selectivity and capacity of visual working memory.

The mechanisms underlying monkey memory have been interpreted assuming monkeys and humans use similar strategies. Our data should temper that assumption, as the tasks used here mimic those used previously. Same–Different mimics “delayed-match-to-color,” used in lesion and electrophysiological studies of short-term memory in monkeys (Bauer and Fuster 1976; Mishkin and Manning 1978). Match First mimics “ABBA” (Miller and Desimone 1994; Miller et al. 1996) or “delayed-match-to-sample” (Pagan et al. 2013), used to contrast selective memorization mechanisms from nonselective detection of stimulus repetition. Match Any mimics “multi-item working memory” (Yakovlev et al. 2005) or “serial probe recognition” (Wright 2007; Basile and Hampton 2010), used to study interference processes and working memory updating processes. The set

sizes for the original tasks just cited range from two to thousands, with delayed-match-to-color using two colors, ABBA using five or six images, and multi-sample working memory or serial probe recognition using six, eight, 16, hundreds, or thousands per testing session. Here we manipulated set size in each of the three tasks. Across three common tasks, each examined under multiple conditions, monkeys consistently relied on different strategies than humans even when overall accuracy was high. Thus, as has been recognized by others (Elmore et al. 2011; Basile and Hampton 2013), task demands and overall performance criteria are not, by themselves, adequate to assess whether monkeys and humans use similar strategies.

Acknowledgments

We thank Adin Horovitz, Kate Kempen, and Michelle Henry for assisting with data collection, and Benjamin Basile and Brian Scott for their critical comments that helped improve the manuscript. The views expressed in this article do not necessarily represent the views of the University of Cape Town, NIMH, NIH, or the U.S. Government. This study was supported by the Intramural Research Program of the National Institute of Mental Health.

References

- Baddeley A. 2010. Working memory. *Curr Biol* **20**: R136–R140.
- Baddeley A. 2012. Working memory: Theories, models, and controversies. *Annu Rev Psychol* **63**: 1–29.
- Basile BM, Hampton RR. 2010. Rhesus monkeys (*Macaca mulatta*) show robust primacy and recency in memory for lists from small, but not large, image sets. *Behav Processes* **83**: 183–190.
- Basile BM, Hampton RR. 2013. Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition* **126**: 391–396.
- Bauer RH, Fuster JM. 1976. Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *J Comp Physiol Psychol* **90**: 293–302.
- Cook RG, Wright AA, Sands SF. 1991. Interstimulus interval and viewing time effects in monkey list memory. *Anim Learn Behav* **19**: 153–163.
- Cowan N. 2008. What are the differences between long-term, short-term, and working memory? *Prog Brain Res* **169**: 323–338.
- Eichenbaum H. 2008. *Learning & memory*, 1st ed. W. W. & Company, Norton.
- Elmore LC, Ma WJ, Magnotti JF, Leising KJ, Passaro AD, Katz JS, Wright AA. 2011. Visual short-term memory compared in rhesus monkeys and humans. *Curr Biol* **21**: 975–979.
- Freeman MF, Tukey JW. 1950. Transformations related to the angular and the square root. *Ann. Math. Statist.* **21**: 607–611.
- Kahana MJ. 2012. *Foundations of human memory*, Oxford University Press, New York.
- Logie RH. 2011. The functional organization and capacity limits of working memory. *Curr Dir Psychol Sci* **20**: 240–245.
- Miller EK, Desimone R. 1994. Parallel neuronal mechanisms for short-term memory. *Science* **263**: 520–522.
- Miller EK, Erickson CA, Desimone R. 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* **16**: 5154–5167.
- Mishkin M, Manning FJ. 1978. Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res* **143**: 313–323.
- Pagan M, Urban LS, Wohl MP, Rust NC. 2013. Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci* **16**: 1132–1139.
- Scott BH, Mishkin M, Yin P. 2012. Monkeys have a limited form of short-term memory in audition. *Proc Natl Acad Sci* **109**: 12237–12241.
- Wickens TD. 2002. *Elementary signal detection theory*, Oxford University Press, New York.
- Wittig JH Jr, Richmond BJ. 2014. Monkeys rely on recency of stimulus repetition when solving short-term memory tasks. *Learn Mem* **21**: 325–333.
- Wright AA. 2007. An experimental analysis of memory processing. *J Exp Anal Behav* **88**: 405–433.
- Yakovlev V, Bernacchia A, Orlov T, Hochstein S, Amit D. 2005. Multi-item working memory—a behavioral study. *Cereb Cortex* **15**: 602–615.
- Zar JH. 2010. *Biostatistical analysis*, 5th ed. Prentice Hall, Upper Saddle River, New Jersey.

Received January 23, 2016; accepted in revised form May 8, 2016.