

On perfect working-memory performance with large numbers of items

Jonathan E. Thiele · Michael S. Pratte ·
Jeffrey N. Rouder

Published online: 1 June 2011
© Psychonomic Society, Inc. 2011

Abstract Many popular models conceptualize working memory as consisting of three or four discrete slots or bins. This conceptualization, however, has been seemingly refuted by Bays and Husain (2009), who reported perfect performance on a working memory task with a large number of very simple items. We show, however, that this perfect-performance result likely reflects a design flaw rather than mnemonic structure. The flaw is that the test array itself in Bays and Husain's study provides information about the correct answer without recourse to working memory. We show perfect performance on eight items for 18 participants when this information is present. We show that performance is poorer, however, when this information is removed. Hence, the Bays and Husain result does not threaten models that stipulate that working memory is composed of limited slots.

Keywords Working memory · Capacity · Working memory capacity

Working memory refers to the information that is consciously available at a point in time. It is well known that such memory is highly limited, but the nature of the limit remains controversial (Cowan & Rouder, 2009; Miyake & Shah, 1999; Osaka, Logie, & D'Esposito, 2007). One popular conceptualization is that working memory consists

of a fixed number of discrete slots in which items or chunks are temporarily held (Cowan, 2001; Luck & Vogel, 1997; Miller, 1956). Nonetheless, there are alternative viewpoints in which fine gradations of mnemonic resources may be spread across several objects (Bays & Husain, 2008; Wilken & Ma, 2004). We refer to the former view as the *discrete-slots model* and to the latter alternative as the *distributed-resources model*.

The evidence for the discrete-slots model comes from three types of results. First, in a variety of paradigms, a model-based estimate of the number of discrete slots converges to about three or four items in adults (Cowan, 2001). Cowan, Fristoe, Elliott, Brunner, and Saults (2006), for example, showed that model-based estimates of the number of slots were constant at 3.5 items when the number of to-be-remembered items was manipulated across a large range. A second source of evidence is that items are seemingly stored in working memory as unitized wholes, without regard to their complexity. Luck and Vogel (1997), for example, manipulated the number of features in each object and found no effect of this type of complexity, supporting the view that objects are stored as wholes. Similar manipulations with cubes and Chinese characters from Awh, Barton, and Vogel (2007) and Barton, Ester, and Awh (2009) also support the invariance of capacity with respect to item complexity. A third source of evidence is that working memory performance seems to follow two distinct patterns: performance either reflects characteristics of the studied item or seems to be from a noninformative guessing state (Rouder, et al., 2008; Zhang & Luck, 2008).

Several studies have questioned these findings (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wheeler & Treisman, 2002; Wilken & Ma, 2004). Perhaps the one study that most forcefully calls into question the

J. E. Thiele · J. N. Rouder (✉)
Department of Psychological Sciences, University of Missouri,
210 McAlester Hall,
Columbia, MO 65211, USA
e-mail: rouderj@missouri.edu

M. S. Pratte
Psychology Department, Vanderbilt University,
301 Wilson Hall,
Nashville, TN 37240, USA

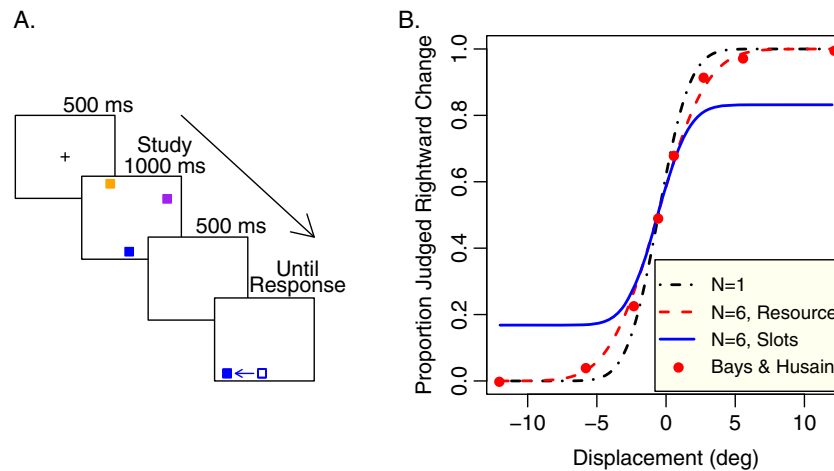


Fig. 1 (A) The paradigm used in the present study is adapted from Bays and Husain (2009). (B) Predictions of working memory models and data from Bays and Husain (2009). Predictions for arrays of one item are identical for the discrete-slots and distributed-resources models (dashed-and-dotted curve). Predictions differ, however, for arrays of six items. The discrete-slots model predicts imperfect

performance, even with large displacements (solid curve). The distributed-resources model, in contrast, predicts perfect performance with arbitrarily large displacements (dashed curve). The data from Bays and Husain (2009) in the six-item condition are denoted as points, and these data are seemingly incompatible with the discrete-slots model

discrete-slots model is Bays and Husain's (2009) demonstration of perfect working memory performance for large numbers of to-be-remembered items. This result is troublesome from a discrete-slots point of view because capacity limits are not flexible and cannot accommodate large numbers of items.

In Bays and Husain's (2009) paradigm, squares of various colors were presented at study. At test, just a half second later, a single square was presented, but it was displaced to either the left or right of its studied position. Figure 1A shows the paradigm we have adapted from Bays and Husain (2009), and it captures the important elements of their design.¹ If one square is presented at study, memory performance follows a very reasonable pattern—performance improves as the displacements become larger in magnitude, and performance quickly approaches ceiling. This behavior is shown in Fig. 1B as the dashed-and-dotted curve. The dependent measure is the proportion of trials judged to have a rightward displacement. This proportion is near .5 when the displacement is small, approaches 0.0 for large leftward displacements and approaches 1.0 for large rightward displacements. The limits of 0.0 and 1.0 denote perfect performance for leftward and rightward displace-

ments, respectively. The critical case is when more items are presented at study than can be stored in working memory. According to the discrete-slots model, some of these items will not be in memory at all, and participants must occasionally guess without information. The consequence of this guessing is that even for arbitrarily large displacements, performance will not be perfect. Predictions for the discrete-slots model are shown by the solid curve in Fig. 1B. A contrasting prediction for this critical case of many items is made by the distributed-resources model. Although each item may receive only a small portion of the fixed mnemonic resources, this small portion is presumably sufficient for detecting arbitrarily large displacements. The consequence is that perfect performance may be obtained, even with large numbers of to-be-remembered items, if the items are sufficiently displaced. This predicted behavior is shown by the dashed curve in Fig. 1B. The key difference is that while the distributed-resources model can account for perfect performance, the discrete-slots model cannot.

Bays and Husain (2009) examined performance for six items and found that it was perfect for large displacements across 4 participants (their data are shown as the circles in Fig. 1B). In fact, this result is replicable, and we show here the same perfect-performance result across a sample of 18 naive participants memorizing eight items. On its face, this perfect-performance result contradicts the discrete-slots model. We believe, however, that the Bays and Husain (2009) paradigm is flawed, and the perfect-performance result is an artifact of this flaw. In this report, we introduce a modified version of the Bays and Husain (2009) paradigm that does not suffer from this flaw. Performance in our version of the task does not reach perfection, even with

¹ There are a few minor procedural differences between the experiment reported here and Bays and Husain (2008, 2009): (1) Bays and Husain presented the study array of squares offset to either the left or right, so that they could measure initial eye movements. We presented our study array centered. (2) Bays and Husain maintained their fixation cross throughout study. We provided a fixation cross first, but removed it during study. (3) Bays and Husain displayed test items for 250 ms, whereas we displayed them until response. Given that we replicated their perfect-performance results, it may be safely concluded that these minor variations are inconsequential.

large displacements. We conclude that the perfect-performance result is not valid and cannot be used to argue against a discrete-slots working memory model.

The flaw in Bays and Husain's (2009) paradigm is that the position of the item at test is correlated with the direction of displacement. Figure 1 demonstrates this correlation. Even without observing the study array, it is obvious that the test item must have moved left because it could not have been further left at study. This correlation is present even when the test item is not so close to the display edge. Most troubling, it grows more pronounced with larger displacements. Participants may adopt a simple strategy: Respond "leftward" if the test item is on the left side of the screen, and "rightward" otherwise. This strategy produces high levels of accuracy for large displacements, even in the absence of any memory for the studied display. In fact, performance from this guessing strategy will become perfect as the displacements grow arbitrarily large. We worry that Bays and Husain's (2009) perfect performance result may reflect a guessing strategy rather than a fundamental signature of the working memory system.

To correct for this flaw, the paradigm may be modified so that the position of the test item reveals no information about the direction of displacement. To do so, we simply presented the test item at the horizontal center of the screen. The remaining study locations were randomly chosen from a distribution symmetric around the center, and consequently, the test display provided no information about the direction of displacement. If perfect performance with large displacements for large numbers of to-be-remembered items is a signature of the mnemonic system, the perfect-performance result should be observed even when the test item is presented at the center of the display.

In the following experiment, we manipulated the testing procedure. In the *test-anywhere* condition, adapted from Bays and Husain (2009), the position of each item at study was drawn randomly. The position of each test item was determined by adding a displacement amount to the studied position. In the *test-at-center* condition, the studied position was chosen such that the position of the test item was horizontally centered on the screen.

Method

Participants A group of 36 students (21 female, 15 male) at the University of Missouri completed the experiment as part of an Introduction to Psychology course requirement.

Design and stimuli The factors of Array Size (either one, three, or eight items), Displacement (10 levels), and Testing Condition (test anywhere vs. test at center) were manipulated in a mixed design. Array size and displacement were

varied in a within-subjects manner; test condition was varied in a between-subjects manner, with 18 participants per test condition level.

In the test-anywhere condition, the position of the studied squares were selected at random from a $12^\circ \times 22.5^\circ$ region centered on the display. One item was displayed at test, and it was displaced from the study position as follows. To ensure that people were not always at ceiling, we used the following displacements: When one item was studied, the 10 levels of displacement were $\{\pm 0.4^\circ, \pm 0.8^\circ, \pm 1.2^\circ, \pm 1.6^\circ, \pm 12^\circ\}$. Displacements for the three-item and eight-item arrays were $\{\pm 0.8^\circ, \pm 1.6^\circ, \pm 2.4^\circ, \pm 3.2^\circ, \pm 12^\circ\}$ and $\{\pm 1.6^\circ, \pm 3.2^\circ, \pm 4.8^\circ, \pm 7.2^\circ, \pm 12^\circ\}$, respectively. Note that in all conditions, the largest absolute displacement was 12° of visual angle.

In the test-at-center condition, the positions of all studied items that were not to be tested were selected at random from a $26.4^\circ \times 15^\circ$ region centered on the display. The tested item was studied at a point such that the test was always at the center (at the negative value of the displacement level). The levels of displacement were $\{\pm 0.6^\circ, \pm 1.2^\circ, \pm 1.8^\circ, \pm 2.4^\circ, \pm 12^\circ\}$ for one-item study arrays, $\{\pm 1.2^\circ, \pm 2.4^\circ, \pm 3.6^\circ, \pm 4.8^\circ, \pm 12^\circ\}$ for three-items study arrays, and $\{\pm 2.4^\circ, \pm 4.8^\circ, \pm 7.2^\circ, \pm 9.6^\circ, \pm 12^\circ\}$ for the eight-item arrays. Note that in all conditions, the largest absolute displacement was 12° of visual angle.

The items in each display were squares that subtended 0.9° of visual angle on a side. Items were drawn at one of either 13 (test-anywhere) or 11 (test-at-center) fixed vertical positions on the screen. Horizontal positions chosen such that the total distance between the centers of any two squares subtended at least 1.75° . Each square on a trial was a unique color, with the colors drawn randomly from a palette of eight primary colors. Participants were seated 50 and 72.5 cm from the screen for the test-anywhere and test-at-center conditions, respectively. The screen itself, a 17-in. Apple iMac display, subtended a $42.4^\circ \times 26.5^\circ$ region in the test-anywhere condition and a $28.8^\circ \times 18^\circ$ region in the test-at-center condition. Displays were presented with the Psychophysics Toolbox (Kleiner, Brainard, & Pelli, 2007).

Procedure Trials began with 500 ms of fixation, followed by the study array presented for 1,000 ms. Following study, a blank screen was presented for 500 ms, followed immediately by the test item, which remained visible until a response was made. Following response, a 1,000-ms blank screen preceded the start of the next trial. Participants indicated that the test item had moved to the left or right by using the "z" and "/" keys with the left and right index fingers, respectively. Participants received positive auditory feedback following correct responses, and no feedback following incorrect ones. Sessions began with a practice block of 6 trials. This was followed by seven experimental blocks of 60 trials each.

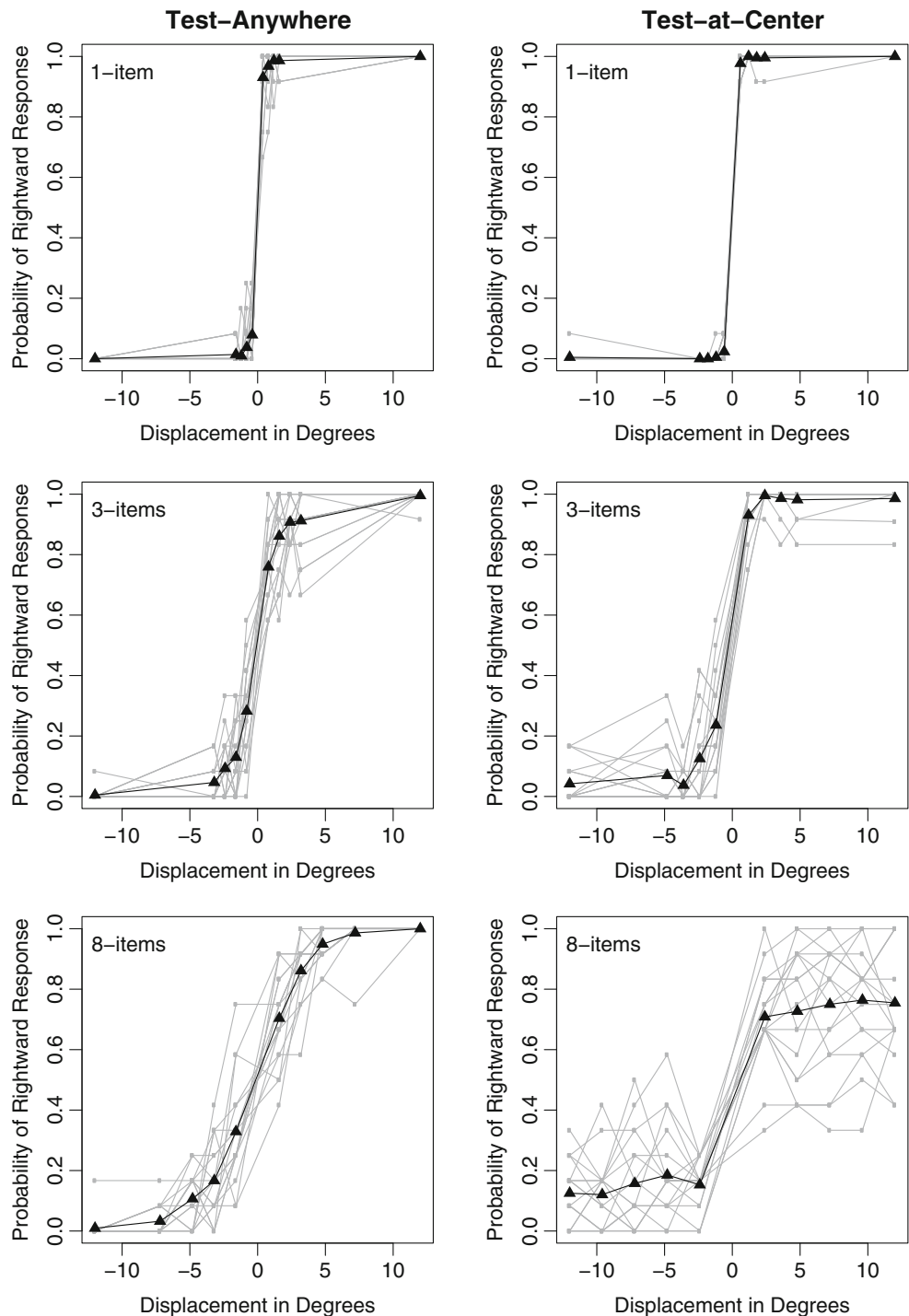
Results

Data from 1 participant in the test-at-center condition were discarded because responses were exceedingly fast. For the remaining participants, responses faster than 200 ms and slower than 10 s were discarded. These trials comprised less than 0.1% and 0.4% of total trials for the test-anywhere and test-at-center conditions, respectively. Other reasonable

choices of response windows, such as from 250 ms to 2 s, did not change the results appreciably.

Figure 2 shows the proportions of rightward responses as a function of displacement across testing and array size conditions. In these plots, the values 0.0 and 1.0 serve as the perfect-performance limits for leftward (negative) and rightward (positive) displacements, respectively. The left and right columns show data from the test-anywhere and

Fig. 2 Proportions of rightward responses as a function of displacement across array size and testing conditions. The gray and black curves denote individual and averaged proportions, respectively. Performance approaches ceiling levels in the test-anywhere condition, but remains error prone in the test-at-center condition



test-at-center conditions, respectively. As can be seen, performance in the test-anywhere condition is nearly perfect for a range of displacements for array sizes of one, three, and even eight items. These findings replicate Bays and Husain's (2009) result that performance can reach ceiling with sufficiently large displacements, even for numbers of items that exceed typical capacity estimates. Performance in the test-at-center condition, displayed on the right in Fig. 2, is perfect or nearly perfect for arrays of one and three items with large displacements. In the eight-item condition, however, average performance never reaches ceiling with large displacements. Instead, performance asymptotes at about 80%, and the vast majority of participants do not achieve near-perfect performance.

Discussion

Bays and Husain (2009) found that participants were able to reach perfect performance for large displacements even after studying arrays of six items. We replicated this result with arrays of eight items across 18 naive participants. Yet this result only held in the test-anywhere condition, the one in which test position was necessarily correlated with the direction of displacement. When this correlation was removed by placing test items in the horizontal middle of the screen, participant performance was reliably less than perfect. Therefore, the observed perfect performance may be due to artifactual correlation rather than to supercapacity working memory performance.

It is reasonable to wonder whether placing the test at the horizontal center changed the nature of the paradigm. In the test-anywhere condition, participants had motivation to encode the position of the item. In the test-at-center condition, however, this motivation was not present. Participants need not remember the actual studied positions, but, rather, whether the item was left or right of center. It seems plausible, therefore, that the position was coarsely and categorically coded, with the categories "left" and "right." While we readily admit that coding may have varied across the testing conditions, this variation does not threaten the lack-of-perfect-performance conclusion. According to a distributed-resources account, memory varies with resources, and if anything, the test-at-center condition with coarse coding would require no more resources than the test-anywhere condition. Consequently, participants should have done as well in the test-at-center condition as they did in the test-anywhere condition. They did markedly worse, however, indicating that the high performance in the test-anywhere condition reflected the methodological design flaw rather than supercapacity working memory performance.

It may seem natural to ask whether the test-at-center data better support a discrete-slots or a distributed-resources

model. We think the paradigm is not especially well suited for this question, for a number of reasons. First, we worry about coarse and categorical coding, which may possibly promote chunking or grouping (Cowan, 2001; Miller, 1956). If participants chunked or grouped, we would not necessarily expect to see a constant capacity. Other testing procedures that discourage chunking, such as Zhang and Luck's (2008), are better candidates for competitively testing the models. Second, the design choices were made to best assess perfect performance. They are not ideal for differentiating between the models. In particular, we do not have enough small displacements with eight items to assess how the slopes of the response functions change across set size.

Conclusion

In this article, we have shown that Bays and Husain's (2009) claim that working memory may be perfect for many items is seemingly the result of an artifact. In their task, there is an unwanted correlation between displacement and test position, with extreme test positions providing information about the direction of displacement without recourse to memory. We contend that participants use this correlation, and have shown that performance is far from perfect when the test position is uncorrelated with displacement direction. Therefore, the perfect-performance result is not a signature of the working memory system and cannot be used to falsify the discrete-slots approach.

Author Note This research was supported by NSF Grant SES-0720229.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.
- Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1359–1367.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.
- Bays, P. M., & Husain, M. (2009). Response to comment on "Dynamic shifts of limited working memory resources in human vision." *Science*, *323*, 877d.
- Cowan, N. (2001). The magic number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–114.
- Cowan, N., Fristoe, N., Elliott, E., Brunner, R., & Sauls, J. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*, 1754–1768.

- Cowan, N., & Rouder, J. N. (2009). Comment on “Dynamic shifts of limited working memory resources in human vision. *Science*, 323, 877c.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception*, 36(ECVP Abstract Supplement).
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
- Osaka, N., Logie, R. H., & D’Esposito, M. (2007). *The cognitive neuroscience of working memory*. Oxford: Oxford University Press.
- Rouder, J.N., Morey, R.D., Cowan, N., Zwilling, C.E., Morey, C.C., & Pratte, M.S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, 105, 5976–5979.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131, 48–64.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120–1135.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235.