Research Article

Psychological Science
24(10) 2038–2047

DOI: 10.1177/0956797613484938
pss.sagepub.com

Two Items Remembered as Precisely as One: How Integral Features Can Improve Visual Working Memory

Gi Yeul Bae and Jonathan I. Flombaum
Johns Hopkins University

Abstract

In the ongoing debate about the efficacy of visual working memory for more than three items, a consensus has emerged that memory precision declines as memory load increases from one to three. Many studies have reported that memory precision seems to be worse for two items than for one. We argue that memory for two items appears less precise than that for one only because two items present observers with a correspondence challenge that does not arise when only one item is stored—the need to relate observations to their corresponding memory representations. In three experiments, we prevented correspondence errors in two-item trials by varying sample items along task-irrelevant but integral (as opposed to separable) dimensions. (Initial experiments with a classic sorting paradigm identified integral feature relationships.) In three memory experiments, our manipulation produced equally precise representations of two items and of one item.

Keywords

visual memory, short-term memory, size discrimination, auditory perception

Received 10/22/12; Revision accepted 3/8/13

Despite its importance for a variety of human behaviors, visual working memory (VWM) seems to be severely limited. Researchers have long hypothesized that an upper bound limits the total number of objects the system can store, sometimes referred to as a "magic number" of approximately four (Cowan, 2001; Luck & Vogel, 1997; Sperling, 1960). Some researchers have questioned the presence of a discrete limit, arguing that VWM is continuously limited by a dynamically distributed resource (Bays & Husain, 2008; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004) that is believed to constrain the resolution with which features are stored. Regardless of the existence of an upper bound on memory capacity, there seems to be consensus that VWM is continuously limited within hypothesized bounds (Brady, Konkle, & Alvarez, 2011; Fukuda, Awh, & Vogel, 2010). Indeed, the severity of VWM limits are perhaps best reflected by the increased difficulty of remembering two things rather than one. Experiments using a wide array of stimuli have shown that each of two items simultaneously held in memory seems less precise than a single item held in memory (Alverez & Cavanagh, 2004; Anderson, Vogel, & Awh, 2011; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Wilken & Ma, 2004; Zhang & Luck, 2008).

In the current study, we challenged the consensus that declines in precision for two items result from competition for limited memory. Instead, we observed that a correspondence challenge hampers memory for two items but not for one. To use a memory to make a judgment about a presently viewed object (e.g., Bays & Husain, 2008; Luck & Vogel, 1997), an observer must determine which stored memories correspond to the observed object. Likewise, to use a memory to report the features of a cued object (e.g., van den Berg et al., 2012; Zhang & Luck, 2008), an observer must determine to which stored memory a cue corresponds. But if only one object is stored, then no such computation is necessary. We suggest that, in fact, the precision of two objects in memory

Corresponding Author:
Jonathan I. Flombaum, Department of Psychological and Brain Sciences, Johns Hopkins University, Ames Hall, 3400 N. Charles St., Baltimore, MD 21218
E-mail: flombaum@jhu.edu

is no worse than the precision of one. Instead, in experiments in which two objects have to be memorized, participants make correspondence errors that have been interpreted as a decline in precision.

If correspondence errors account for apparent costs associated with remembering two items, then costs should be eliminated when correspondence errors are prevented (Bae & Flombaum, 2012; Bae, Wilson, Holland, & Flombaum, 2013). To test this prediction, we sought a method that would support observers when they make correspondence decisions. Logically, observers should use stable but task-irrelevant features of objects to make correspondence decisions, because these are the features that usually do not change between a memory display and a test display. Correspondence decisions should be anchored to stable features of the world. But in most experiments, task-irrelevant features are either identical among all memory items (e.g., identical shapes when color is the target) or different along separable dimensions (e.g., shape and color). Identical features provide no basis for correspondence decisions, whereas separable dimensions provide a problematic basis for decision making because of known feature-binding challenges (Treisman & Schmidt, 1982). Our goal was to design memory displays in which sample items differed along a stable feature dimension that could provide a sound basis for correspondence judgments.

To achieve this end, we found inspiration in prior research focusing on a distinction between integral and separable feature combinations (Garner, 1974; Garner & Felfoldy, 1970). Integral features (e.g., size and shape) are those that can be manipulated independently but are not perceived independently. Separable features (e.g., size and color) are represented independently, and binding between them is not always successful (Treisman & Schmidt, 1982). We hypothesized that distinguishing items via separable features (as is often done) leaves the correspondence problem in place. For example, if people are asked to remember the sizes of two triangles that also differ in color (see Experiment 4), the presence of a color at test will not necessarily help with the correspondence problem; the problem of determining which remembered size corresponds to, for example, "the blue one" is no less difficult than determining which remembered size corresponds to "the one over there." In contrast, a uniquely identifying integral feature should help. In an experiment on size memory, when presented with a circle as a probe, for example, one would naturally query the sizes of remembered circles only, ignoring remembered triangles. Including only a single circle in a memory set of two could, in turn, largely prevent correspondence errors. Following this logic, we hypothesized that integral feature distinctions would erase costs associated with remembering two items rather than one item.

We tested this hypothesis in memory experiments for three different features: the luminance of two objects when they possessed the same hue (e.g., both were a version of red) or different hues (e.g., one was red and one green); the sizes of two objects when they were either the same shape or different shapes; and, in an auditory experiment, the amplitudes of two tones when they were either the same frequency or different frequencies. We predicted equal performance for two items compared with one item when the two items in memory had different integral features. In addition, to validate the specific integral features used, we conducted a conceptual replication of classic sorting experiments by Garner and Felfoldy (1970).

Experiments 1 and 2: Sorting Integral Features

To independently identify integral feature dimensions in vision, we conceptually replicated a now classic experiment by Garner and Felfoldy (1970). In Experiment 1, three colored squares with different luminance values were presented simultaneously, and participants were asked to find the brightest and second brightest squares (Fig. 1a). The three colored squares could be from the same hue family or from different hue families. We predicted that performance with items sharing a hue would be better than with items from different hue families, because hue and luminance are integral features. This was the logic underlying Garner's experiments (Garner, 1974; Garner & Felfoldy, 1970). In Experiment 2, we applied the same logic to size and shape. Participants were instructed to identify the smallest and then the second smallest items, by area, in a set of three (Fig. 1b). The items were either all the same shape or different shapes. We expected worse performance for the sets containing different shapes than for the sets containing the same shape.

Method

Participants. Two groups, each containing 10 undergraduates at Johns Hopkins University, participated in exchange for course credit. One group participated in Experiment 1 (luminance sorting), and the other participated in Experiment 2 (size sorting). All participants had normal or corrected-to-normal visual acuity. The protocols for these experiments were approved by the Johns Hopkins University Institutional Review Board.

Apparatus, stimuli, and procedure. Participants sat in a darkened room 60 cm from an iMac display subtending approximately 39.56 × 25.35 degrees of visual angle (horizontal × vertical). Stimuli used in the experiments are shown in Figure 1. The stimuli for Experiment 1

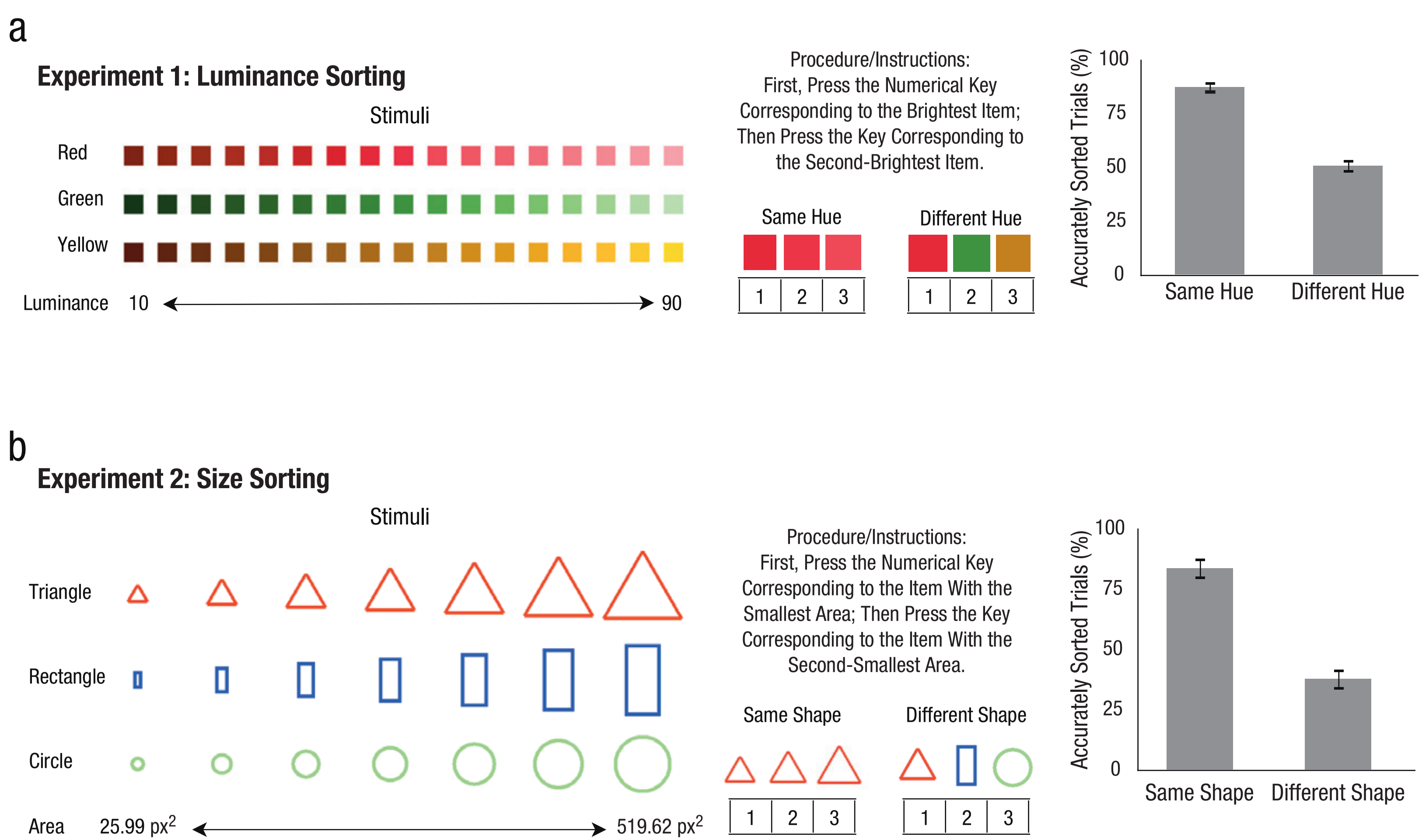


Fig. 1. Experimental design and results for (a) Experiment 1 and (b) Experiment 2. In Experiment 1, stimuli were created from a set of 51 colored squares (17 each of red, green, and yellow hues). Each hue family was generated with the same 17 luminance values ranging between 10 and 90, in steps of 5 (stimuli in each column shown here share the same luminance value). On each trial, squares in three different colors were selected from the set of 51 colors and presented in a random order horizontally at the center of a black screen. In half of the trials, the colors were selected from the same hue family; in the remainder of the trials, one item from each of the three hue families was selected. The task was to sort the presented items from brightest to darkest by pressing keys corresponding to the numbers under each item. (Numbers shown in the figure were not presented in the display, but participants learned during practice that 1 through 3 designated left to right, respectively.) Identical procedures were used in Experiment 2, except that there were three different shapes (objects in each column shown here were of equal area). The hue and luminance of these shapes remained constant, but the area they subtended varied. In half of the trials, the shapes were the same; in the remainder of the trials, one item from each of the three shape families was selected. The task was to sort the shapes from smallest to largest on the basis of area by pressing keys as in Experiment 1. The graphs show the mean percentage of trials on which stimuli were correctly sorted as a function of trial type. Error bars represent ±1 *SEM*.

consisted of a set of 51 colored squares (17 with red hues, 17 with green hues, and 17 with yellow hues). Hue families were defined by a and b coordinates in Commission internationale de l'éclairage (CIE) L*a*b* color space (red: a = 80, b = 38; green: a = −40, b = 38; yellow: a = 20, b = 98). Each hue family was generated with the same 17 possible luminance (L) values ranging between 10 and 90, in steps of 5. The stimuli for Experiment 2 consisted of a set of 21 colored shapes (7 triangles, 7 rectangles, and 7 circles) with consistent luminance and hue (red, blue, and green, respectively) but with each shape in its group subtending a different area (range = 25.99–519.62 pixels^2).

On each luminance-sorting trial, we chose three colored squares from the set of 51 possibilities (Fig. 1a). In half of the trials, all the colors presented were from the same hue family. In the other half, each item was a different hue. In all trials, the three squares varied in luminance (i.e., we varied the L coordinate in CIE L*a*b* color space). Squares were positioned horizontally in a random order. The task was to pick the brightest and then the second brightest squares by pressing numerical keys corresponding to the squares' positions on the screen (left, center, and right corresponding to 1, 2, and 3, respectively).

An identical procedure was used in the size-sorting experiment (Fig. 1b). Three objects with either the same shape or different shapes were presented. The task was to pick the item with the smallest area and then the item with the second smallest area by pressing numerical keys corresponding to the shapes' positions on the screen (left, center, and right corresponding to 1, 2, and 3, respectively). Each experiment consisted of 120 trials and 10 practice trials.

Results and discussion

Sorting performance was worse when the items were from different hue families than when they were from the same hue families, $t(9) = 8.90$, $p < .05$, $d = 3.04$ (Fig. 1a), and worse when they were different shapes than when they were the same shapes, $t(9) = 12.48$, $p < .05$, $d = 3.73$ (Fig. 1b). These results illustrate what is perhaps a cardinal rule of perception: that it is often more than the sum of its parts. Features that are independently controllable may not combine independently. Consequently, it can be challenging for observers to compare objects along one dimension when they are different along a second, seemingly irrelevant but integral dimension. We exploited this difficulty in three additional experiments.

Experiments 3 to 5: Preventing Memory Errors With Integral Features

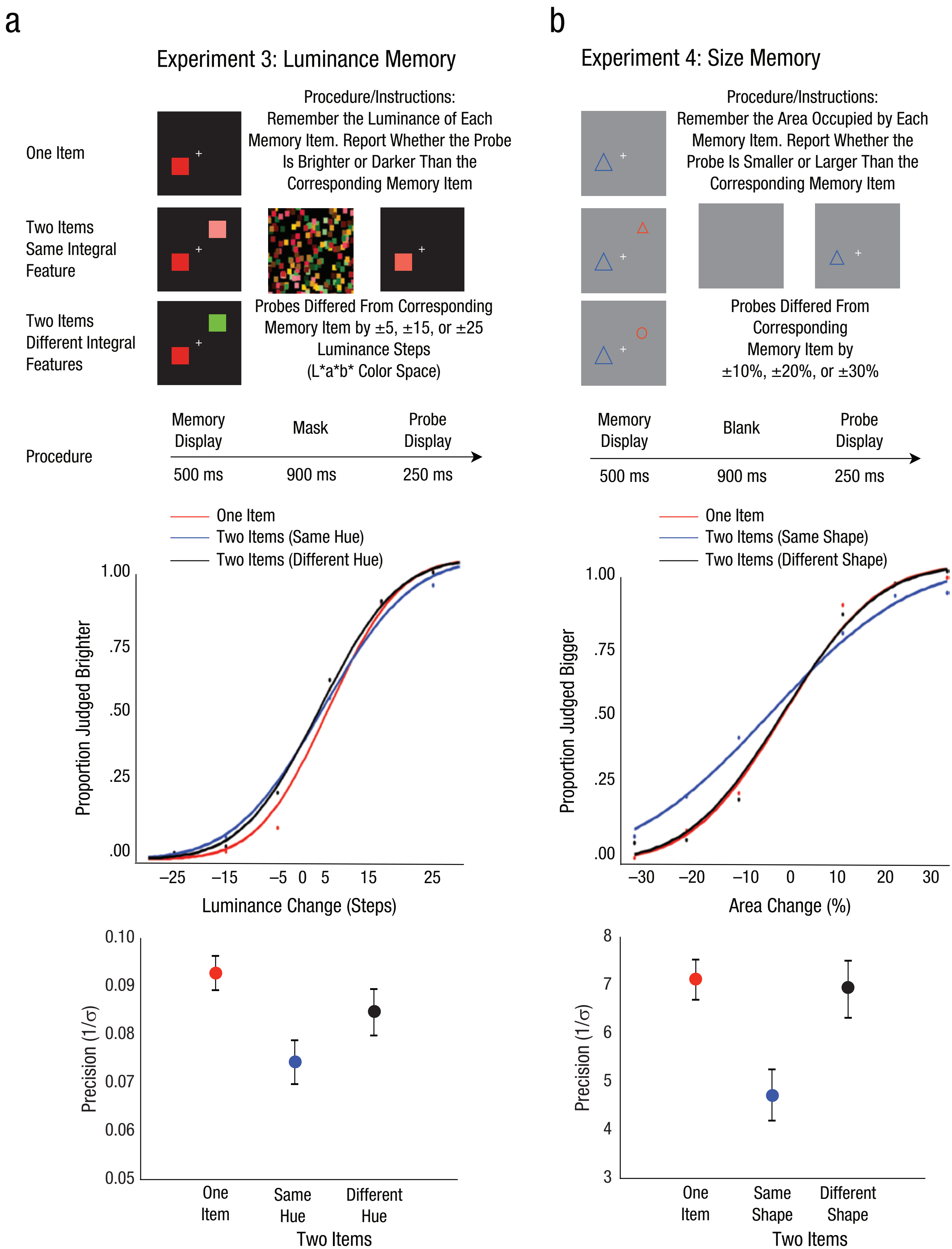
Armed with specific knowledge about dimensions that are integral, and with the knowledge that differences in one integral dimension can make comparisons along another dimension challenging, we sought to prevent correspondence problems in a VWM experiment. We used a basic feature-judgment task (e.g., Bays & Husain, 2008), in which an observer is instructed to remember, for example, the luminance of either one or two items. At test, a single probe item appears, and the observer must compare its luminance with that of the corresponding memory item. In two-item trials, a correspondence error can take place for items that share the same hue family. The observer might compare the probe to the wrong memory item, potentially producing an erroneous response (and worse performance in two-item trials compared with one-item trials). In contrast, what might happen if the two objects differ along an integral dimension? Though observers may generally make correspondence errors, Experiments 1 and 2 showed that mistaken correspondences should, in this situation, produce difficult and unintuitive comparisons; for example, which is brighter: the green one I see or the red one I remember? The unintuitive nature of the comparison could potentially lead an observer to realize (not necessarily explicitly) that he or she has made a correspondence mistake, and the observer might then address the task comparison to the correct memory item instead.

Thus, we predicted that two items would be remembered as precisely as one when the two items had different, but task-irrelevant, integral features. When two items shared the same feature, however, we predicted typical costs for two compared with one. On the basis of the results of Experiments 1 and 2, we tested these predictions. In Experiment 3, we tested memory for luminance with objects that differed or were the same in hue, and in Experiment 4, we tested memory for size with objects that were different shapes or the same shape. In addition, we extended our manipulation to auditory working memory in Experiment 5 by testing memory for loudness using tones with the same or different frequencies (amplitude and frequency being known integral auditory features; Garner, 1976; Wood, 1975).

Method

Participants. A new group of Johns Hopkins University undergraduates participated. There were 12 participants in Experiment 3, 8 in Experiment 4, and 12 in Experiment 5. All participants had normal or corrected-to-normal vision and hearing. The protocols for these experiments were approved by the Johns Hopkins University Institutional Review Board.

Apparatus, stimuli, and procedure. In Experiments 3 and 4, the apparatus and stimuli were identical to those used in Experiments 1 and 2, respectively. The luminance experiment (Experiment 3; Fig. 2a), consisted of one- and two-item trials. In each trial, one or two squares were


a
Experiment 3: Luminance Memory
Procedure/Instructions:
Remember the Luminance of Each Memory Item. Report Whether the Probe Is Brighter or Darker Than the Corresponding Memory Item
One Item
Two Items Same Integral Feature
Two Items Different Integral Features
Probes Differed From Corresponding Memory Item by ±5, ±15, or ±25 Luminance Steps (L*a*b* Color Space)
Procedure
Memory Display
Mask
Probe Display
500 ms
900 ms
250 ms
One Item
Two Items (Same Hue)
Two Items (Different Hue)
Proportion Judged Brighter
1.00
.75
.50
.25
.00
–25
–15
–5
0
5
15
25
Luminance Change (Steps)
Precision (1/σ)
0.10
0.09
0.08
0.07
0.06
0.05
One Item
Same Hue
Different Hue
Two Items
b
Experiment 4: Size Memory
Procedure/Instructions:
Remember the Area Occupied by Each Memory Item. Report Whether the Probe Is Smaller or Larger Than the Corresponding Memory Item
Probes Differed From Corresponding Memory Item by ±10%, ±20%, or ±30%
Memory Display
Blank
Probe Display
500 ms
900 ms
250 ms
One Item
Two Items (Same Shape)
Two Items (Different Shape)
Proportion Judged Bigger
1.00
.75
.50
.25
.00
–30
–20
–10
0
10
20
30
Area Change (%)
Precision (1/σ)
8
7
6
5
4
3
One Item
Same Shape
Different Shape
Two Items

(continued)

Fig. 2. (continued)

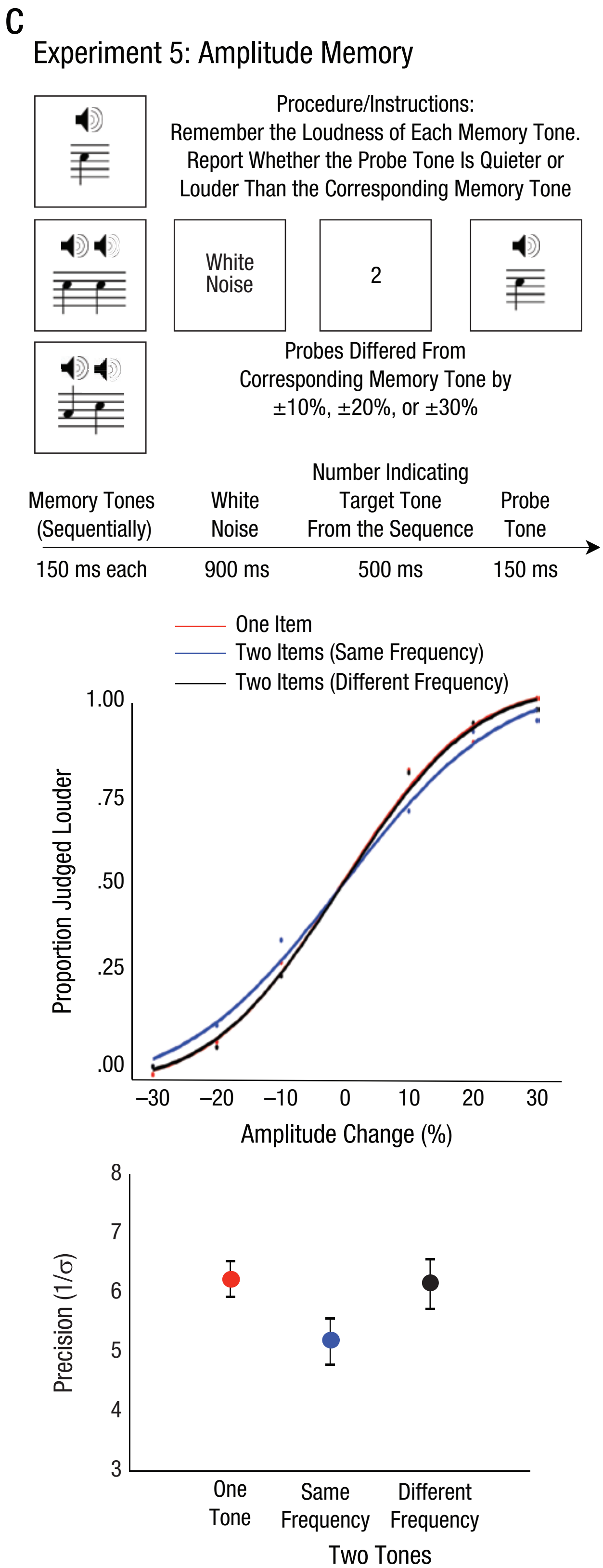


Fig. 2. Experimental procedures (top row), estimated response functions (middle row), and estimated precisions (bottom row) for Experiments 3, 4, and 5 (a, b, and c, respectively). Stimuli consisted of squares (Experiment 3), circles and triangles (Experiment 4), and auditory tones (Experiment 5). In each experiment, there were three trial types: one item only, two items with the same integral feature, and two items with different integral features. The integral feature that was manipulated in Experiment 3 was hue; in Experiment 4, it was shape; and in Experiment 5, it was the frequency of the tones. In each experiment, the stimulus display was followed by a mask and then a probe item. Participants were asked to indicate whether the probe was brighter or darker (Experiment 3), smaller or larger (Experiment 4), or quieter or louder (Experiment 5) than the corresponding memory item. Response functions for each trial type were estimated in terms of the probability of a "brighter," "larger," or "louder" response as a function of the magnitude of change in luminance, area, or amplitude (in Experiments 3, 4, and 5, respectively). Reciprocals of standard deviations of estimated functions are plotted as a function of trial type, separately for the three experiments. Error bars represent ±1 *SE*.

presented in random positions around a central fixation cross. Each square was assigned 1 of 17 luminance values and a hue drawn from the 51 shown in Figure 1a. There were two kinds of two-item trials. In the same-integral-feature trials, both squares had the same hue. In different-integral-feature trials, each square had a different hue. Participants were asked to memorize the luminance of each item. After a brief mask display, a probe item appeared in the same position and with the same hue as one of the corresponding memory items. Participants judged whether the probe was darker or brighter than the memory item. A probe differed from its corresponding memory item by ±5, ±15, or ±25 L-value steps. There were 432 trials plus 10 practice trials.

The size experiment (Experiment 4; Fig. 2b) was nearly identical to the luminance experiment. Memory objects were circles and triangles. One-item trials included either a circle or a triangle. Two-item trials included either two objects with the same shape (same-integral-feature trials) or two objects with different shapes (different-integral-feature trials). The objects in both trial types were always different colors (red and blue). The task was to remember the area occupied by each item. At test, a probe item appeared at the same position and with the same shape and color as the corresponding memory item. The size of a probe was ±10%, ±20%, or ±30% relative to the memory item. Participants reported whether the probe was larger or smaller than its corresponding memory item. There were 216 trials plus 10 practice trials.

In Experiment 5 (Fig. 2c), auditory stimuli were delivered in stereo through headphones. On each trial, one or two pure tones of either 220 or 400 Hz were played. In two-item trials, the tones were separated by a 500-ms silent interval and a 150-ms burst of white noise. The amplitudes of the tones were chosen randomly from a range of 57.1 to 64.6 dB, with a minimum difference of 1.25 dB. Participants were instructed to remember the loudness of each tone. After a second burst of white noise (in two-item trials), a *1* or *2* was displayed to indicate which tone from the sequence was to be compared with the upcoming probe. The probe tone always had the same frequency as its corresponding memory tone, but it varied in amplitude by ±10%, ±20%, or ±30%. Participants judged whether a probe's amplitude was softer or louder than that of the memory tone. In same-integral-feature trials, the two memory tones had the same frequency, and in different-integral-feature trials, they had different frequencies. There were 216 trials plus 10 practice trials.

Analysis. To compare the quality of memory representations, we fit psychometric functions with a probit regression model in each memory load and condition (Fig. 2). The reciprocal of the standard deviation of these functions (Fig. 2) reflects the precision of memory representations (Bays & Husain, 2008; Palmer, 1990).

Results and discussion

The graphs in Figure 2 plot the response functions obtained from each trial type (one item, two items with the same integral features, and two items with different integral features) and the estimates of representational precision derived from these functions. There were significant effects of trial type in each experiment—luminance: $\chi^2(2, N = 5{,}184) = 15.8$, $p < .001$, $d = 3.97$; size: $\chi^2(2, N = 1{,}728) = 26.4$, $p < .001$, $d = 5.13$; and amplitude: $\chi^2(2, N = 2{,}592) = 8.0$, $p = .018$, $d = 2.83$. Planned comparisons explored differences between one-item trials and the two kinds of two-item trials. A typical cost arose for remembering two items compared with one when the two items shared an integral feature value. These costs can be seen in the statistically significant difference between the estimated standard deviations of the response functions for the different trial types in the luminance experiment, $z = -3.892$, size experiment, $z = -4.357$, and amplitude experiment, $z = -2.478$ (ps < .05). Note that in the size experiment, costs were present despite the fact that the two shapes always differed in color, which is separable from shape.

In contrast, costs were eliminated for two items compared with one when the two items included different integral features—luminance experiment: $z = -1.62$, size experiment: $z = 0.32$, and amplitude experiment: $z = -0.165$ (ps > .1). Representational precision was also significantly better in different-integral-feature trials than in same-integral-feature trials—luminance experiment: $\chi^2(1, N = 3{,}456) = 5.5$, $p = .019$, $d = 2.34$, size experiment: $\chi^2(1, N = 1{,}152) = 16.6$, $p < .001$, $d = 4.07$, and amplitude experiment: $\chi^2(1, N = 1{,}728) = 5.4$, $p = .02$, $d = 2.32$.

We also fit models that included a guessing parameter, a standard feature of many psychophysical models, incorporating the assumption that some responses are not made on the basis of a probe-versus-stimulus comparison but as unbiased guesses instead. The main effects with respect to representational precision were the same (see Additional Analyses: Model With Guessing and Table S1 in the Supplemental Material available online). In addition, we explored potential edge-driven effects in the paradigm (Thiele, Pratte, & Rouder, 2011). Such effects were present in Experiments 3 and 5 (but not in Experiment 4), though the magnitude of any such effects did not vary across trial types (see Additional Analyses: Edge Effects in the Supplemental Material).

Finally, we sought positive evidence of correspondence errors. We sorted trials into two categories on the basis of the response to the probe demanded by the target and nontarget memory items. *Consistent* trials were those in which the probe differed from each memory item in the same way (e.g., the probe was darker than both of the memory items). *Inconsistent* trials were those in which the probe differed from each memory item in different ways (e.g., the probe was darker than the target memory items but lighter than the nontarget memory item). Correspondence errors should lead inconsistent trials to show a larger error rate than consistent trials. We found a significant difference in this direction in the same-integral-feature trials of Experiments 3 and 5 and a marginal difference in the same-integral-feature trials of Experiment 4 (Figs. S1 and S2 in the Supplemental Material). Specifically, in Experiment 3, the mean error rates were 44% and 9% for inconsistent and consistent trials, respectively, $t(11) = 14.199$, $p < .01$; in Experiment 4, the mean error rates were 26% and 16% for inconsistent and consistent trials, respectively, $t(7) = 2.143$, $p = .06$; and in Experiment 5, the mean error rates were 37% and 15% for inconsistent and consistent trials, respectively, $t(11) = 4.04$, $p < .01$. We also compared error rates in inconsistent trials across same- and different-integral-feature trials to obtain evidence that integral features reduced the frequency of errors. There was a significant difference in error rates in Experiment 3 (inconsistent trials: $M = 44\%$, consistent trials: $M = 37\%$), $t(11) = 2.95$, $p = .013$, and in Experiment 4 (inconsistent trials: $M = 26\%$, consistent trials: $M = 7\%$), $t(7) = 3.94$, $p < .01$, but not in Experiment 5 (inconsistent trials: $M = 37\%$, consistent trials: $M = 32\%$), $t(11) = 0.64$, n.s. (although the effect in the latter was in the right direction).

These results are consistent with the prediction that preventing correspondence errors eliminates VWM costs for two objects compared with one. In the typical case—that is, in same-integral-feature trials—standard deviations of estimated response functions are thought to reflect only representational precision. But performance in different-integral-feature trials suggests that they may also reflect correspondence errors for which our comparison of consistent and inconsistent trials provides further evidence.

General Discussion

In the experiments reported here, we observed that when two items shared the same integral feature, there was an apparent and typical decline in memory quality for those items compared with memory quality for one item. But when the two memory items possessed different integral features, we did not observe declines in performance.

This is, to our knowledge, the only one of many recent studies to eliminate entirely a memory cost for two items compared with one (Bays & Husain, 2008; van den Berg et al., 2012; Zhang & Luck, 2008; but see Bae et al., 2013). The presence of such a cost is a critical prediction of both flexible-resource theories (Bays & Husain, 2008; van den Berg et al., 2012) and hybrid-resource theories (Alvarez & Cavanagh, 2004; Anderson et al., 2011), in which the quality of two items in memory should always be roughly half that of a single item. (Of current theories, only a traditional fixed-capacity model would not make this prediction; e.g., Cowan, 2001; Luck & Vogel, 1997). It is important that although the memory items had different integral features, the participants did not know which would be probed, and the same feature of each item had to be remembered. Thus, the effects cannot be explained in terms of different pools of resources dedicated to the storage of different features.

Instead, any account must address the role of integral features in eliminating memory costs. We have supplied one such account, exploiting the fact that two-item trials usually present observers with a correspondence problem, whereas one-item trials do not. We hypothesized that differences in integral features would prevent attendant correspondence errors by supplying observers with reliable and salient anchors for correspondence decisions. In general, any theory that acknowledges noisy or probabilistic representations of object features must also acknowledge that correspondence computations are necessary for retrieving a memory, although the mechanisms underlying these computations are rarely discussed (but see Bae et al., 2013; Levillain & Flombaum, 2012).

There is a second potential account of our results, also in terms of correspondence errors, but during perception as opposed to test. Specifically, researchers often conceive of perception and encoding as the noisy sampling of features from images (Girshick, Landy, & Simoncelli, 2011; Vul, Hanus, & Kanwisher, 2009; Vul & Rich, 2010). During perception or encoding, then, there also exists a correspondence challenge when more than one item is present. After drawing a sample with some feature content, an observer needs to assign a correspondence between the feature and one of the objects believed to be in the image. Because samples are noisy—because an observer should be uncertain about where exactly in time and space a sample came from—there is a risk of correspondence errors. Indeed, this exact explanation has recently been offered to account for well-known feature-binding challenges (Treisman & Schmidt, 1982; Vul & Rich, 2010). Of course, no feature-binding challenge exists for one-item displays, so typical costs associated with remembering two items compared with one may reflect encoding correspondence errors (i.e., instances during which stray samples influence the inferred properties of an object). And, in turn, differences in integral features may prevent such sampling-related errors by making it clearer when pairs of samples arose from independent sources.

It may seem incompatible with previous evidence that apparent changes in precision for one and two items are actually driven by correspondence errors, particularly because several models have incorporated a "misbinding" parameter meant to capture these instances (Bays et al., 2009; but see Anderson et al., 2011). However, all modeling thus far has assumed random causes of correspondence errors, an equal likelihood in all trials regardless of particular stimulus properties and arrangements. If correspondence errors have systematic causes—if they are more likely in some trials than in others—then current models may not estimate their prevalence correctly. For example, correspondence errors may be more likely the nearer that two items are to one another (Emrich & Ferber, 2012; Vul & Rich, 2010). Moreover, our results suggest that surface similarity may play a role. If items are more likely to be confused when they share an integral feature than when they do not, perhaps the continuous extent of any similarity modulates the likelihood of correspondence error. This possibility should be explored in future research.

Explaining our results in terms of how integral features prevent correspondence errors requires neural mechanisms that represent integral feature combinations. Recent functional MRI research suggests a potential mechanism: Neurons can be tuned to conjunctions of features. Integral features, then, are those for which neurons reflect joint preferences, whereas separable features are those for which pairs of neurons represent conjunctions (Drucker, Kerr, & Aguirre, 2009). Joint tuning could supply a basis for making correspondence decisions that use integral feature differences.

Finally, what is perhaps most surprising about the results reported here is that memory performance improved when memory displays were made more complicated. In Experiment 4, memory performance was better when displays included a triangle and a circle, for example, as opposed to two triangles, yet one could more efficiently summarize the contents of two-triangle displays. Redundancy in these scenes should have recruited less memory resources than the varied displays. We therefore accounted for the performance observed not in terms of memory resources and storage, but in terms of correspondence computations that must be involved in encoding and retrieving contents to and from memory. In general, research has focused almost exclusively on the nature of VWM resources and storage, to the exclusion of the computations that must be involved in acquiring and using information. As we have shown here, accounting for such computations—and the errors that they may induce—can lead to the realization that storage limitations are less severe than they may initially seem. VWM resources seem at least ample enough to afford representations of two items that are as precise as representations of just one. But computations deciding between multiple options naturally become more error prone as the number of options increases (see also, Duncan, 1980; Navon, 1984).

Author Contributions

All authors developed the experimental concept and contributed to the design. Testing and data collection were performed by G. Y. Bae. G. Y. Bae performed the data analysis and interpretation under the supervision of J. I. Flombaum. G. Y. Bae drafted an early version of the manuscript, and then J. I. Flombaum and G. Y. Bae edited the manuscript together. All authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

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.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a plateau when individual item-limits are exceeded. *Journal of Neuroscience*, *31*, 1128–1138.

Bae, G. Y., & Flombaum, J. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*, 703–715.

Bae, G. Y., Wilson, C., Holland, P. C., & Flombaum, J. (2013). *Correspondence computations limit spatial working memory*. Manuscript submitted for publication.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), Article 7. Retrieved from http://www.journalofvision.org/content/9/10/7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision*, *11*(5), Article 4. Retrieved from http://www.journalofvision.org/content/11/5/4

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Drucker, D. M., Kerr, W. T., & Aguirre, G. K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology*, *101*, 3310–3324.

Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology*, *12*, 75–96.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4), Article 12. Retrieved from http://www.journalofvision.org/content/12/4/12

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.

Garner, W. R. (1974). *The processing of information and structure*. New York, NY: Wiley.

Garner, W. R. (1976). Interaction of stimulus dimensions in concepts and choice processes. *Cognitive Psychology*, *8*, 98–123.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225–241.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*, 926–932.

Levillain, F., & Flombaum, J. (2012). Correspondence problems cause repositioning costs in visual working memory. *Visual Cognition*, *20*, 669–695.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, *91*, 216–234.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 332–350.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1–29.

Thiele, J. E., Pratte, M. S., & Rouder, J. N. (2011). On perfect working-memory performance with large numbers of items. *Psychonomic Bulletin & Review*, *18*, 958–963.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA*, *109*, 8780–8785.

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, *138*, 546–560.

Vul, E., & Rich, A. N. (2010). Independent sampling of features enables conscious perception of bound objects. *Psychological Science*, *21*, 1168–1175.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), Article 11. Retrieved from http://www.journalofvision.org/content/4/12/11

Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, *104*, 3–20.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.