The transition from feature to object: Storage unit in visual working memory depends on task difficulty

Jiehui Qian¹ · Ke Zhang¹ · Shengxi Liu¹ · Quan Lei²

Published online: 2 July 2019
© The Psychonomic Society, Inc. 2019

Abstract

Visual working memory (VWM) is a cognitive memory buffer for temporarily processing and storing visual information. Previous studies suggest that its capacity is severely limited, and there is an ongoing debate on whether the storage capacity is object-based or feature-based in VWM. In this study, a change-detection task was employed to investigate whether and how task difficulty can affect VWM, specifically, its capacity and the unit of storage. Task difficulty was manipulated through the set size of memory items, memory fidelity required by the resolution of representation and the type of feature tested. We examined two types of stimuli: the single-feature type, where each memory item was composed of a single feature (color or shape), and the conjunctive-feature type, where each item was composed of a conjunction of two features (colored shape). Experiment 1 replicated the previous findings that memory capacity for colors was larger than shapes, and decreased with the resolution demand regardless of the type of stimuli. In Experiment 2, we analyzed and compared the results from single-feature items and conjunctive-feature items in the low- and high-resolution conditions while controlling for the number of to-be-remembered features. By directly matching the estimated capacity based on an object-unit and a feature-unit with the theoretical prediction, the results showed that the unit storage in VWM tended to be feature-based with low task difficulty, and to be object-based with high task difficulty. This suggests that VWM is dynamic and flexible, dependent on the load of the current task.

Keywords Visual working memory · Task difficulty · Feature-based storage · Object-based storage

Introduction

Although research on verbal working memory is extensive (for review, see Baddeley, 2003), there has been increasing attention over the last two decades on how information is retained in visual working memory (VWM). VWM is responsible for temporarily holding, processing, and manipulating visual information. It helps to maintain perceptual stability across discontinuation and variations in retinal image as a result of eye, head, or body movement, and is

also suggested to correlate with fluid intelligence (Fukuda, Vogel, Mayr, & Awh, 2010; Unsworth, Fukuda, Awh, & Vogel, 2015). It provides a crucial link between perception and other cognitive processes, therefore investigating the mechanisms underlying VWM may be essential for understanding vital brain functions.

Previous studies show that VWM is severely limited in capacity, with only 3–4 items capable of being stored from a given display (Alvarez & Cavanagh 2004; Awh, Barton, & Vogel, 2007; Cowan 2001; Luck & Vogel 1997; Vogel, Woodman, & Luck, 2001; Zhang & Luck 2008). In Luck and Vogel (1997) classic study on VWM, the authors suggested that the unit of storage in VWM is object rather than feature, a theory known as the 'slot' model (Zhang & Luck, 2008). They employed a change-detection paradigm that required observers to compare a briefly presented memory array and a following test array, then to indicate whether there had been a change between the two. They found that multi-feature objects can be retained in working memory as well as single-feature objects. Consistent with their findings, some studies show that VWM capacity is confined by the number of objects and is independent of the number

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-019-00956-y>) contains supplementary material, which is available to authorized users.

✉ Jiehui Qian
qianjh3@mail.sysu.edu.cn

¹ Department of Psychology, Sun Yat-Sen University, Guangzhou, China

² Department of Psychology, University of Minnesota, Minneapolis, MN, USA

of features within each object (Fukuda et al., 2010; Vogel et al., 2001; Xu, 2002a). However, there is research showing that the fidelity of memory, i.e., the precision of the memory representations, may decrease with objects' featural complexity while the quantity of stored objects can be largely unaffected by increasing number of features within an object (Magnussen & Greenlee 1999; Fougny, Asplund, & Marois, 2010). Some researchers attributed the deteriorated memory performance on complex stimuli to a more difficult comparison process during the change detection task (Awth et al. 2007; Jiang, Shim, & Makovski, 2008; Zhang & Luck 2008). This is because complex stimuli generally have higher similarity within a specific category, e.g., Hebrew characters, resulting in a more difficult and erroneous comparison process. Furthermore, electrophysiological evidence shows that the contralateral delay activity (CDA)—a neural indicator of VWM capacity—is determined by the number of stored items but not of the features (Luria & Vogel 2011; McCollough, Machizawa, & Vogel, 2007; Vogel & Machizawa 2004; Wilson, Adamo, Barense, & Ferber, 2012).

By contrast, a number of researchers have reported worse memory performance and lower object-based capacity (Alvarez & Cavanagh, 2004; Delvenne & Bruyer, 2004; Olson & Jiang, 2002; Wheeler & Treisman, 2002) for more complex objects or feature–feature conjunctions (Bays, Catalao, & Husain, 2009; Bays & Husain 2008; Bays, Wu, & Husain, 2011; Eng, Chen, & Jiang, 2005; Gao, Li, Liang, Chen, Yin, & Shen, 2009; Olson & Jiang 2002; Ma, Husain, & Bays, 2014), suggesting that the storage mechanism in VWM is not solely object-based and feature may be the storage unit as well. For the feature-based hypothesis, it is the number of features that limits the capacity of VWM. Evidence shows that the memory performance for color–color conjunctions is worse than that for single colors (Delvenne & Bruyer, 2004; Olson & Jiang, 2002; Wheeler & Treisman, 2002), but the performance does not seem to worsen for conjunctions like color–shape (Wheeler & Treisman, 2002), size–orientation (Olson & Jiang, 2002), and shape–texture (Delvenne & Bruyer, 2004). To account for the divergence among these mixed results, researchers proposed several models that exclude either object or feature as the sole unit of storage in VWM, e.g., 'weak object-based' hypothesis (Olson & Jiang, 2002; Vergauwe & Cowan, 2015), 'Boolean map' hypothesis (Huang & Pashler 2007; Shen, Yu, Xu, & Gao, 2013), etc. Specifically, Wheeler and Treisman (2002) suggested that features from the same dimension compete for capacity, whereas features from different dimensions can be stored in parallel. In other words, an object-based unit can be used for storing conjunctions consisted of features from different dimensions and a feature-based unit for storing same-dimension feature conjunctions. Although this theory seems to explain

many of the inconsistent findings, there is still evidence showing that the memory performance for multi-dimension feature conjunctions is even worse than the most difficult single features (Parra, Sala, Logie, & Morcom, 2014) and the debate on whether a consistent object benefit exists for multi-dimension feature conjunctions continues.

Studies investigating the question of storage unit in VWM often employ a change-detection paradigm to compare the memory performance for a single-feature display and a multi-feature display with either equal number of objects or features (Lee & Chun, 2001; Olson & Jiang, 2002; Xu, 2002a). There are two main predictions that could be drawn from a hypothesis of an object-based unit: (1) remembering a multi-feature object should be easier than remembering the same number of features on multiple objects (or locations), and (2) encoding a task-relevant feature of an object should automatically involves encoding the other task-irrelevant features of the object. Although there is evidence showing no or little cost when presenting features over a larger set of objects or locations (Fougny et al., 2010), studies that support an 'object benefit' often found a decrease in change detection-performance when the same number of features are distributed over a larger number of objects (Delvenne & Bruyer 2004; Fougny, Cormiea, & Alvarez, 2013; Xu 2002b). The decrease in memory performance may be due to the difficulty of filtering out the task-irrelevant features in mixed displays and leaking of memory resources to the task-irrelevant features, however a wider distribution of spatial attention over more objects/locations during the encoding or maintenance may also confound the results (Shin & Ma, 2017). Indeed, Treisman's Feature Integration Theory (FIT; Treisman & Gelade, 1980) states that selective attention may allow separate features to be linked via their shared spatial location and further bound into an object. Although it seems that storing features of an object relies on attending to a specific location, it is controversial whether the spatial factor plays a crucial role in the effect of 'object benefit'. A few studies suggested that the 'object benefit' may in fact reflect a location-based benefit since there are fewer associated locations to remember for feature conjunctions than for equivalent number of separate features (Wang, Cao, Theeuwes, Olivers, & Wang, 2016), while others showed that the limiting factor of VWM capacity is the number of objects rather than the number of spatial locations of relevant visual stimuli (Lee & Chun, 2001).

The question of whether the storage in VWM is object-based or feature-based seems to rely upon many factors, e.g., the type of stimuli tested (Delvenne & Bruyer, 2004; Olson & Jiang, 2002; Wheeler & Treisman, 2002), the testing situation (Vergauwe & Cowan, 2015), etc. Based on the past literature, here we propose that task difficulty, as a common factor underlying these seemingly vastly different

factors, could account for the divergence of the existing findings. Task difficulty reflects the cognitive demand to accurately encode and store a memory display. Factors influencing task difficulty can be coarsely characterized into two categories: the type of task performed and the nature of visual stimuli that a task performed on. The former can be determined by factors like whether to memorize feature conjunctions or a single feature, whether to include an additional memory load task, or the type of feature required to be stored, etc. The latter can be determined by factors like the number of memory items presented (set size), the complexity or visual features of a memory item, etc. Specifically, task difficulty differs among: (1) different set size, e.g., a set size of six memory items is more difficult to memorize than a set size of one and evidence consistently shows that memory performance decreases with set size (Luck & Vogel, 1997; Zhang & Luck, 2008); (2) different feature types, e.g., memorizing colors is easier than shapes since color is considered perceptually more salient and resolvable than shape (Gegenfurtner, 2003; Thornton & Gilden, 2007), and evidence consistently shows that memory performance for colors is better than for shapes; (3) different feature values in the same feature type, e.g., different color categories (such as red vs. green) are more discriminable than different shades of the same color (such as dark red vs. light red), because the latter requires a finer resolution of memory representation and thus task difficulty increases (Gao, Ding, Yang, Liang, & Shui, 2013; Olson & Jiang 2002). Although studies investigating the above factors are abundant, so far there is no consensus on how these factors affect the storage process and the unit of storage in VWM.

In the present study, we varied the task difficulty to investigate its effect on VWM, aiming to advance our understanding of the storage mechanisms involved in VWM. The aforementioned three factors that may influence task difficulty, set size, feature type, and the required resolution of representation, were tested using a change-detection paradigm. The required resolution of representation was manipulated by varying the degree to which a memory item and the probe differed. Since a smaller difference between the two requires a more precise storage for the memory item, the resolution demand and thus the task difficulty increases. In order to investigate how visual information is being stored in VWM, memory capacity was used as a probe to look into the unit of storage in VWM. Experiment 1 replicated an effect that the memory capacity decreases with the resolution demand (Gao et al., 2013) and compared the hit rates between different testing conditions to evaluate the validity of an object-/feature-based storage unit in VWM. In Experiment 2, we investigated how the unit of storage in VWM is affected by task difficulty by directly matching

the estimated capacity based on an object-unit and on a feature-unit with the theoretical prediction.

Experiment 1: effect of feature type and resolution on VWM for single- and conjunctive-feature stimuli

Experiment 1 was aimed at replicating the previously reported findings that the capacity estimates were different for various types of stimuli (Alvarez & Cavanagh, 2004) and it declined as the required resolution of representation increased (Gao et al. 2009, 2013; Gao, Gao, Li, Sun, & Shen, 2011; Ye, Zhang, Liu, Li, & Liu, 2014). We tested the single-feature stimuli, colors and shapes, as well as the conjunctive-feature stimuli, colored shapes. In addition, the hit rates were compared between the single-feature and conjunctive-feature conditions to assess the possible storage unit in VWM.

Methods

Participants Fifteen undergraduates from Sun Yat-Sen University (SYSU) with normal or corrected vision participated in the Experiment 1 for payment. The sample size was chosen based on two criteria: (1) previous literature and (2) a power analysis. Previous work has used a range of sample size from 5 to 24 to look at the effects similar to ours, with a typical effect size of $\eta^2 = 0.85$ (Vergauwe & Cowan, 2015). We conducted a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), which revealed that at least ten participants would be required to have an 90% power to detect the effect in our study. Participants were all naive to the purpose of this study. Written informed consent approved by SYSU Institutional Review Board (IRB) was obtained from each participant before the experiments.

Materials The stimuli were displayed on a uniform black background and viewed on a 23-inch HP proDisplay P231 monitor, with a resolution of 1600 × 900 and a refresh rate of 60 Hz. The viewing distance was 70 cm. Participants were seated in a dark room to complete all the experiments.

Two types of stimuli were tested. In the single-feature condition, the memory array consisted of either colors (in the shape of square) or geometric shapes (in gray); in the conjunctive-feature condition, the memory array consisted of colored geometric shapes, i.e., conjunctions of the two single features. The colors were randomly chosen without replacement from a set of seven highly discriminable color categories: red, green, blue, yellow, magenta, cyan, and gray (see Fig. 1). Each category contained three similar colors

Fig. 1 Stimuli and experimental procedures. *Top*: two types of features (color and shape) used in the experiments. Colors consisted of seven categories: red, green, blue, magenta, yellow, cyan, and gray, with each category containing three similar colors that differed in relative brightness. Shapes consisted of five categories: circle, rectangle, pentagon, triangle, and diamond, with each category containing three

shapes that differed in aspect ratio. *Bottom*: the experimental procedures used in the experiments (*left*) and an exemplar of memory array with four features in Experiment 2 (*right*). Each item in single-feature condition was composed of either color (*in square*) or shape (*in gray*); each item in conjunctive-feature condition was composed of both color and shape (color-shape conjunction)

that differed in relative luminance, yielding a total of 21 colors. For example, the ‘red’ category included three colors with RGB values of [85,0,0], [170,0,0], and [255,0,0]. We manipulated luminance of the colors instead of hue since our preliminary experiment showed that individual differences regarding task difficulty were much greater for detecting changes in hue than that in luminance. Each colored square subtended $0.65^\circ \times 0.65^\circ$. Similarly, the geometric shapes were chosen without replacement from a set of five highly discriminable shape categories: circle, triangle, rectangle, diamond, and pentagon (see Fig. 1). Each category contained three similar shapes that differed in aspect ratio, yielding a total of 15 shapes. For example, the ‘circle’ category included a circle, a horizontally elongated oval and a vertically elongated oval. All the single-feature shapes and the conjunctive-feature colored shapes had approximately equal size, subtending about $1.3^\circ \times 1.3^\circ$. The set size tested for colored squares was 3, 4, or 6, for gray shapes was 2, 3, or 4, and for conjunctions was 2, 3, or 4. The selection of set sizes was based on previously reported VWM capacities for color, shape, and conjunctions (Alvarez & Cavanagh 2004; Olson & Jiang 2002; Qian, Zhang, Lei, Han, & Li, 2019).

The stimuli were presented in pseudorandom positions within a 3×2 grid subtending $9.9^\circ \times 7.5^\circ$. The memory items were randomly distributed among the six cells, with

a separation of no less than 2° between any two adjacent items.

Procedure Each trial began with a fixation cross ‘+’ presented at the center of the screen for 500 ms. The memory array was then presented for 250 ms, followed by a 900-ms retention interval in which the display was blank except for the fixation cross. During the test phase, a probe was presented at one of the locations in the memory array. Participants were asked to judge whether the probe was the same as the memory item. If ‘no change’ was detected, participants needed to press ‘F’ on the keyboard; otherwise, ‘T’ on the keyboard. The probe remained on the screen until response. A diagram of task sequence is shown in Fig. 1.

In the single-feature condition, each participant received 216 trials for each type of stimuli (color/shape), with 108 trials for each resolution condition. For each condition, the probe was identical to the memory item in half of the trials. If the probe differed from the memory item: in the low-resolution condition, the changed feature was selected at random from the remaining feature category that had not been used in the memory array (e.g., memory item: red square, probe: green square); in the high-resolution condition, it was selected from the same feature category as the memory item (e.g., memory item: dark red square, probe item: light red square). Participants were informed

beforehand that the changes between the memory item and the probe were large and easy to detect in the low-resolution condition, and that the changes were small and more difficult to detect in the high-resolution condition. The single-feature colors, shapes, and conjunctions, with the different resolutions, were tested in separate blocks. The order was counterbalanced among the participants. In the conjunctive-feature condition, because either color or shape might change in each trial, each participant received 108 trials per resolution condition per type of change (color/shape). Either feature of the probe was selected in the same manner as in the single-feature condition, and only one feature changed in half of the trials. Each participant received a total of 432 trials for the conjunctive-feature condition. Before the experiment, participants performed a practice section with 15 trials to get acquainted with the stimuli and the task. They were required to achieve an accuracy of above 70% in order to continue with the formal experiment.

Data analysis In our analysis, memory capacity referred to the maximum number of individual representations with certain precision that one was capable of holding in memory. We calculated a commonly used memory index, Cowan's K, for estimating the capacity (Cowan, 2001). Cowan's K, which equals $Set\ size \times (Hit\ rates -$

$False\ alarm\ rates)$, was developed for a change detection task as in Luck and Vogel (1997) and was a modification of one presented earlier by Pashler (1988). The formula assumed that the probability of apprehending a certain item equaled the ratio of the maximum number of items allowed (capacity) and the number of items to be remembered (set size), so that capacity could be estimated from hit and false alarm rates given a set size. However, since the K measure confounds detection sensitivity and response bias, caution should be taken while comparing Ks across different conditions (see Pazzaglia, Dube, & Rotello, 2013, for review). Based on the hit rates and false alarm rates (Table 1, and see [Supplementary Materials](#) for ROC curves), our results showed that the response bias was relatively neutral and consistent across conditions, confirming the validity of the K measure in the analysis.

In addition, we compared the hit rates between the single-feature and conjunctive-feature conditions to assess the possible storage unit in VWM. If items in the memory array were stored using an object-based unit, single-feature stimuli and conjunctive-feature stimuli should be memorized equally well. In particular, the hit rates for detecting color (shape) changes should be equal between the single-feature colors (shapes) and the conjunctive-feature stimuli given the same set size. On the other hand, if items in the memory array were stored using a feature-based

Table 1 Hit and false alarm rates (standard errors) of the data

	Hit rates				FA rates			
	Color		Shape		Color		Shape	
	Low	High	Low	High	Low	High	Low	High
Experiment 1								
Single feature								
Set size 3	94 (2)	64 (4)	78 (4)	54 (4)	21 (4)	25 (4)	25 (5)	24(3)
Set size 4	87 (3)	56 (3)	67 (4)	51 (2)	25 (4)	33 (4)	33 (5)	31(5)
Conjunctive feature								
Set size 3	87 (3)	59 (4)	69 (4)	60 (3)	23 (4)	31 (4)	23 (4)	31 (4)
Set size 4	80 (3)	57 (4)	65 (4)	54 (3)	35 (4)	43 (4)	35 (4)	43 (4)
Experiment 2								
Single feature								
Set size 2	98 (1)	78 (4)	88 (3)	70 (3)	6 (2)	16 (3)	8 (2)	20 (2)
Set size 4	95 (1)	66 (4)	70 (4)	67 (4)	14 (3)	33 (5)	25 (4)	23 (3)
Set size 6	87 (4)	68 (3)	63 (4)	56 (5)	22 (5)	42 (4)	29 (4)	36 (6)
Conjunctive feature								
Set size 2	97 (1)	79 (4)	92 (3)	86 (2)	4 (1)	11 (1)	4 (1)	11(2)
Set size 4	95 (1)	69 (4)	83 (3)	74 (3)	11 (3)	22 (3)	11 (3)	22 (3)
Set size 6	90 (2)	73 (3)	79 (4)	69 (4)	23 (3)	32 (3)	23 (3)	32 (3)

Note: "Color" indicates detecting color changes for single-feature stimuli or for conjunctive-feature stimuli; "Shape" indicates detecting shape changes for single-feature stimuli or for conjunctive-feature stimuli

unit, single-feature stimuli should outperform conjunctive-feature stimuli as more features were presented in the latter given the same set size. In other words, the hit rates for detecting color (shape) changes should be higher for the single-feature colors (shapes) than for the conjunctive-feature stimuli given the same set size. By matching the results of hit rates comparisons with the above theoretical predictions, we may be able to test how information is stored in VWM. We did not analyze the false alarm rates or accuracies, since we could not tell whether the false alarms were made on ‘color’ or ‘shape’ for the conjunctive-feature stimuli on the ‘no change’ trials and therefore no proper predictions could be made on the comparisons for either false alarm rates or accuracies.

Results

To examine whether the effect of resolution demand on VWM was different for single-feature colors, shapes, and color-shape conjunctions, we conducted a $3 \times 2 \times 2$ (stimulus type $\times$ resolution $\times$ set size) repeated measures analysis of variance (ANOVA). The data of detecting color changes and shape changes were averaged for the conjunctive-feature condition. Since different set sizes were employed for different types of stimuli, only the common set sizes of 3 and 4 were used in the analysis. We found that stimulus type had a significant effect on capacity,

$F(2, 28) = 14.29$, $p < 0.001$, $\eta_p^2 = 0.51$. Post hoc comparison showed that the capacity was significantly higher for colors than for shapes ($p = 0.01$) and color-shape conjunctions ($p = 0.01$), whereas the latter two were not significantly different from each other ($p = 0.53$). The main effect of resolution was significant, $F(1, 14) = 70.77$, $p < 0.001$, $\eta_p^2 = 0.84$. No main effect of set size was found, $F(1, 14) = 1.68$, $p = 0.20$, $\eta_p^2 = 0.11$. There was a significant interaction between stimulus type and resolution on capacity, $F(2, 28) = 16.56$, $p < 0.001$, $\eta_p^2 = 0.64$. The difference in capacity between the low- and the high-resolution conditions for colors ($\Delta K = 1.7$) was greater than that for shapes ($\Delta K = 0.6$) and color-shape conjunctions ($\Delta K = 0.9$). Overall, the capacity for single-feature items was larger than that for conjunctive-feature in the low-resolution condition but equivalent to conjunctive-feature in the high-resolution condition. No other interaction effect was found ($ps > 0.23$). The results are shown in Fig. 2a (capacity estimates were averaged across set sizes 3 and 4 since there was no significant effect of set size on capacity).

For single features, we further conducted a 2×3 (resolution $\times$ set size) repeated-measures ANOVA separately for each type of stimuli. The results are shown in Fig. 2b for colors and Fig. 2c for shapes. Consistent with the above analysis, the effect of resolution was significant both for the color stimuli, $F(1, 14) = 41.62$, $p < 0.001$,

Fig. 2 Results of Experiment 1. Bar plots show the comparisons between the mean estimated capacities (data averaged across set size 3 and 4) at two levels of resolutions for colors, shapes, and color–shape conjunctions **a**, for colors of set sizes of 3, 4, and 6 **b**, for shapes of set sizes of 2, 3, and 4 **c**, and for color–shape conjunctions **d**. The error bar stands for the standard error of the mean

$\eta_p^2 = 0.75$ and for the shape stimuli, $F(1, 14) = 37.72$, $p < 0.001$, $\eta_p^2 = 0.72$. There was no significant difference across set sizes for either colors ($p = 0.45$) or shapes ($p = 0.78$). No significant interaction effect was found ($ps = 0.09$). For conjunctions, we further compared the estimated capacity between the tasks of detecting color changes and shape changes using a $2 \times 3 \times 2$ (resolution $\times$ set size $\times$ tasks) repeated measure ANOVA. There was a significant effect of resolution on capacity, $F(1, 14) = 63.78$, $p < 0.001$, $\eta_p^2 = 0.82$. An interaction between resolution and task was found on capacity, $F(2, 28) = 21.72$, $p < 0.001$, $\eta_p^2 = 0.61$. In the low-resolution condition, estimated capacity for detecting color changes was significantly higher than that for detecting shape changes ($p = 0.008$), while they were not significantly different in the high-resolution condition ($p = 0.37$). No other interaction effect was found ($ps > 0.18$). The results are shown in Fig. 2d (capacity estimates were averaged across set sizes).

In addition, the results of hit rates comparisons between the conjunctive-feature and single-feature conditions are shown in Table 2. Since we were interested in whether the hit rates for the single-features were higher than conjunctions for each set size and resolution condition rather than the main effect of set size or resolution, paired-sample t tests (one-tailed) were performed instead of repeated-measures ANOVAs. We used an adjustment of false discovery rate (FDR) to correct for an inflated type I error caused by using the multiple paired-sample t tests. The results showed that the hit rates for the conjunctive-feature stimuli were significantly lower than that for the single-feature colors under these conditions: (1) for all tested set sizes when detecting low-resolution color changes; (2) for a

set size of 3 when detecting low-resolution shape changes; (3) for the averaged low-resolution hit rates with a set size of 3. As we mentioned in the *Data analysis* section, a significant decrease in the hit rates for the conjunctive-feature stimuli indicated a feature-based storage unit in VWM. No other significant difference was found.

Experiment 2: effect of task difficulty on the unit of storage in VWM

Experiment 1 showed that VWM performance differed among different types of stimuli and with different resolution demands. This effect has been demonstrated in a number of studies (Alvarez & Cavanagh, 2004; Gao et al., 2013; Ye et al., 2014), despite that the magnitude of the effect may vary with the experimental settings employed in the study (Olson & Jiang, 2002). Additionally, the analysis on the hit rates indicated that the storage unit in VWM varied for different types of feature, resolution demand, and set size, which all reflected variations in task difficulty. It is possible that task difficulty plays an important role in understanding the mechanisms of visual working memory—not only it determines the amount of mental effort one takes to perform a memory task, but also affects the quantity and fidelity of representation in VWM. Depending on task difficulty, one might adopt different strategies to optimize the performance in a memory task, which might further affect the way that visual information is stored in the VWM. Therefore, in Experiment 2, we manipulated the factors that influence task difficulty to directly investigate the unit

Table 2 The results of paired-sample t tests for comparison between the hit rates for detecting changes in conjunctive- and single-feature stimuli

	Color		Shape		Average	
	Low	High	Low	High	Low	High
Experiment 1						
3 objects	$t(14) = -2.31$ $p = .043$	$t(14) = -1.21$ $p = .295$	$t(14) = -2.63$ $p = .024$	$t(14) = .27$ $p = .950$	$t(14) = -2.77$ $p = .018$	$t(14) = -1.01$ $p = .396$
4 objects	$t(14) = -2.25$ $p = .049$	$t(14) = .18$ $p > .999$	$t(14) = -.58$ $p = .685$	$t(14) = 2.09$ $p = .057$	$t(14) = -1.00$ $p = .400$	$t(14) = 1.88$ $p = .097$
Experiment 2						
2 features	$t(21) = -1.02$ $p = .572$	$t(21) = 1.55$ $p = .246$	$t(21) = 2.01$ $p = .100$	$t(21) = 6.76$ $p < .001$	$t(21) = 2.02$ $p = .102$	$t(21) = 4.76$ $p < .001$
4 features	$t(21) = .91$ $p = .671$	$t(21) = 1.12$ $p = .489$	$t(21) = 2.89$ $p = .016$	$t(21) = 3.64$ $p = .003$	$t(14) = 2.65$ $p = .027$	$t(21) = 3.31$ $p = .005$
6 features	$t(21) = .975$ $p = .615$	$t(21) = .303$ $p > .999$	$t(21) = 3.26$ $p = .007$	$t(21) = 3.19$ $p = .009$	$t(21) = 2.92$ $p = .016$	$t(21) = 2.50$ $p = .039$

Note: “Color” indicates the results for detecting color changes; “Shape” indicates the results for detecting shape changes; “Average” indicates the results for averaging the hit rates for detecting color and shape changes

of storage in VWM by employing a regression analysis. The logic of Experiment 2 was that if the same amount of to-be-stored visual information was displayed and the same task was performed, the estimated memory capacity should be the same. We used a similar set of stimuli as that in Fournie et al. (2010), where both feature types (color and shape) were included in each display under the single-feature and the conjunctive-feature conditions. By quantitatively comparing the estimated capacity for single- and conjunctive-feature stimuli under different experimental conditions while controlling for the number of to-be-remembered features, it was possible to reveal how the task difficulty affects the unit of storage in VWM.

Past research suggested that the storage unit in VWM depends on the testing situation and the guidance from instructions (Olson and Jiang, 2002; Vergauwe & Cowan, 2015). However, these studies did not specify under what situation a choice of storage unit is made in absence of explicit guidance from instructions. This experiment aimed to provide a systematic explanation for variations in the storage unit, i.e., linking the storage unit from object-based to feature-based with a crucial factor—task difficulty.

Methods

Participants Twenty-two undergraduates from SYSU with normal or corrected vision participated in the experiment for payment. We recruited more participants in this experiment than in Experiment 1, since in principle regression analyses required higher precision and therefore more participants to obtain accurate estimates on the slopes. Participants were all naive to the purpose of the study. Written informed consent approved by SYSU IRB was obtained from each participant before the experiment.

Materials Two types of stimuli were used in this experiment: single-feature items and conjunctive-feature items. In the single-feature condition, each item in the memory array was composed of one single feature relevant to the task. That is, either the color or shape of an item was required to be memorized and might be tested later. The color was randomly selected from one of the seven color categories except the gray category; the shape was randomly selected from one of the five shape categories except the rectangle category. The shape of square was reserved for the single-feature color stimuli, thus all colors were in the shape of a square; whereas the gray color was reserved for the single-feature shape stimuli, thus all shapes were in gray (Fig. 1). For each item, only one feature might change—a squared item could only change its color whereas a gray item could only change its shape—hence the ‘single-feature’ condition. There were always an equal number of to-be-remembered colors and shapes in the memory array.

A set size of 2, 4, or 6 features (objects) was tested. In the conjunctive-feature condition, each item was a conjunction of two features, i.e., colored shapes, where each feature was selected independently from their respective categories (gray and square were excluded). A same set size of 2, 4, or 6 features (1, 2, or 3 objects) was tested. Since each single-feature item contained only one relevant feature whereas each conjunctive-feature item contained two relevant features, the relevant visual information carried by N objects in the latter was equivalent to that carried by $2N$ objects in the former. The size of each item subtended $1.3^\circ \times 1.3^\circ$. The resolution demand was also manipulated in the experiment, and the observers were informed of the experimental setting beforehand. The apparatus and other settings were the same as in Experiment 1.

Procedure The experimental procedure was the same as in Experiment 1, except for the following changes. For single-feature stimuli, one of the colors (squared items) would change in one fourth of the trials and one of the shapes (gray items) would change in another fourth of the trials. For the conjunctive-feature stimuli, either color or shape of one item would change in half of the trials. Participants were informed of the above information to ensure that they only attended to color for the squared items and to shape for the gray items in the single-feature condition, while they needed to attend to both features in the conjunctive-feature condition. Seventy-two trials were carried out for each resolution condition per stimulus type per set size, yielding a total of 864 trials. The two resolution conditions and the two stimulus types were tested in separate experimental blocks. Sixty practice trials were given for observers to get acquainted with the experimental setting.

Data analysis Because memory capacity was used as a probe to look into the storage unit of VWM in this experiment, we must first clarify the formula used for estimating the capacity. As we mentioned in the *Data Analysis* section of Experiment 1, since Cowan’s K is originally devised for indexing memory capacity in the procedure employed by Luck and Vogel (1997), it also preposes an assumption advocated by Luck and Vogel, that object is the unit of VWM and therefore the *set size* in the formula equals the number of items presented in the display (Cowan, 2001). In this case, the K measure estimates memory capacity in the form of the quantity of individual objects able to be stored (as in Experiment 1). However, previous studies also suggest that feature might be the unit of storage. To accommodate this change, we applied a small modification to the original formula, by letting the probability of apprehending a certain feature be the ratio of the maximum number of features one capable of holding (capacity) and the number of features to be stored (set

size). In the same way as the original formula was derived, memory capacity in the form of the quantity of individual features could be estimated from hit and false alarm rates given a set size of features, i.e., the *set size* in the formula should equal the number of features to be stored. Therefore, the definition of *set size* in the formula needs to be specified according to the proper storage unit in VWM.

Ideally, one's memory capacity should remain constant if the same amount of visual information is provided for memorizing and the same memory task is performed. In this experiment, the same amount of visual information (in terms of feature load) was provided when the number of to-be-membered colors and shapes was equaled in the single- and conjunctive-feature conditions, therefore the estimated VWM capacity should be the same under these two conditions if a proper formula was used. In other words, a mismatch in capacity estimates between the single- and conjunctive-feature conditions may reflect a misuse of the formula—an improper selection of the storage unit. Thus, by comparing the degree of equivalence in the estimated capacity between the two conditions using the different storage units could assess whether an object is stored as an integrated whole or as its constituent features. Note here, the K value of estimated capacity might be different given a different storage unit. Since conjunctive-feature stimuli contained two relevant features, K that estimated based on a feature unit would be twice as that estimated based on an object unit. On the other hand, single-feature stimuli contained only one relevant feature, K would be the same using either a feature-based unit or an object-based unit. However, this does not mean that the actual amount of visual information stored in the brain changes with the storage unit, it simply reflects a difference in the unit of measurement, just as there is the same amount of orange juice whether or not it is served in one big bottle or three small cups. In addition, the absolute value of capacity is not of our primary interest, we focused on how the pattern of variations in the estimated capacity changes with task difficulty to demonstrate whether the storage unit in VWM is fixed or flexible.

For data analysis, since the hit and false alarm rates (Table 1, and see [Supplementary Materials](#) for ROC curves) showed that the response bias was relatively neutral and consistent across conditions, the use of the K measure was valid. Capacity estimates for individual observers were calculated using a feature-based unit and an object-based unit, respectively, for each experimental condition (stimuli type, set size, resolution, and changed-feature type). Individuals' capacity estimates with the same unit were matched between the single-feature and the conjunctive-feature stimuli. To quantitatively assess the relationship between the capacity estimates for single-feature stimuli (K_s) and for conjunctive-feature stimuli (K_c), a

linear regression analysis was performed for each resolution condition, set size, and changed-feature type (in a preliminary analysis, a logarithmic fit was also assessed but its goodness of fit was no better than that for a linear fit). The intercept of the fitted line was set to zero, indicating a theoretical extreme case when one has no VWM capacity. Ideally, if the unit of storage was selected appropriately, the slope of the fitted line should equal to 1, i.e., capacity estimates for single-feature (K_s) and for conjunctive-feature (K_c) stimuli should be equal when their visual information load was the same. Therefore, the closeness of the fit to the theoretical prediction line, $K_c = K_s$, indicated the properness of a specific storage unit. Specifically, using a feature-based capacity estimation, $K_{c/f} = K_{s/f}$ indicated a valid feature-based unit, whereas using an object-based estimation, $K_{c/o} = K_{s/o}$ indicated a valid object-based unit. In our experiment, since $K_{s/f} = K_{s/o}$ and $K_{c/f} = 2K_{c/o}$, we could further obtain that: using a feature-based estimation, $K_{c/f} = 2K_{s/f}$ suggested an object-based unit; and vice versa, using an object-based estimation, $K_{c/o} = \frac{1}{2}K_{s/o}$ suggested a feature-based unit. Here, we provided 95% confidence intervals (CIs) of the slopes to compare with 1 and to indicate the closeness between the fit and 1. In addition, Bayes factors (BFs) were provided to indicate the credibility of the slopes equaling 1. For Bayes factor analysis, the null hypothesis (H_0) was $slope \neq 1$ and the alternative hypothesis (H_1) was $slope = 1$, therefore $BF_{10} > 1$ indicates that evidence favors H_1 .

In addition to the regression analysis, we also compared the hit rates between the single-feature and conjunctive-feature conditions to assess the possible storage unit in VWM. The logic of the theoretical predictions was the same as in Experiment 1, except that in Experiment 2 comparisons were made between the single- and conjunctive-feature stimuli that had the equal number of relevant features (instead of equal number of objects in Experiment 1). If items in the memory array were stored using a feature-based unit, single-feature stimuli and conjunctive-feature stimuli should be memorized equally well, and their hit rates should be equal. On the other hand, if items in the memory array were stored using an object-based unit, conjunctive-feature stimuli should outperform single-feature stimuli as more items were presented in the latter. In other words, the hit rates should be better for the conjunctive-feature stimuli than that for the single-feature stimuli.

Results

Figure 3 shows $K_{c/f}$ (y-axis) and $K_{s/f}$ (x-axis) for each observer with the filled symbols, and Fig. 4 shows $K_{c/o}$ (y-axis) and $K_{s/o}$ (x-axis) for each observer with the hollow symbols. For each resolution condition, set size,

Fig. 3 Relationship between capacity estimates for single-feature (x -axis) and conjunctive-feature (y -axis) stimuli, assuming a feature-based unit. The *top two panels* show the averaged data for detecting color and shape changes, the *middle panels* show the data for detecting color changes, and the *lower panels* show the data for detecting shape changes. The *left and right panels* show the data for the low- and high-resolution conditions, respectively. Each *datum* indicates an observer, and each *line* shows the least-square linear fit to the data (slopes shown in legend). Set size is indicated by both the color and shape of the symbol

Fig. 4 Relationship between capacity estimates for single-feature (x -axis) and conjunctive-feature (y -axis) stimuli, assuming an object-based unit. The *top two panels* show the averaged data for detecting color and shape changes, the *middle panels* show the data for detecting color changes, and the *lower panels* show the data for detecting shape changes. *Left and right panels* show the data for the low- and high-resolution conditions, respectively. Each *datum* indicates an observer, and each *line* shows the least-square linear fit to the data (slopes shown in legend). Set size is indicated by both the color and shape of the symbol

and changed-feature type, the best linear fits (with zero intercept) of the data using a feature-based unit are plotted in Fig. 3 by solid lines and that using an object-based unit are plotted in Fig. 4 by dashed lines. Their slopes, 95% CIs, r^2 s, and BF_{10} s are listed in Table 3.

Overall, the two figures show that: 1) the slope of the line increased with set size in each panel; 2) the slope increased as the resolution demand increased, comparing the left panels with the right panels; 3) the slope also increased for shape (bottom panels) compared to color (middle panels). Since the set size, resolution and feature type all influenced task difficulty, the results indicated that the slope of the fitted line increased with task difficulty. Specifically, the slope using a feature-based data set deviated farther from 1 while that using an object-based data set gradually approached 1. This indicates that VWM tends to be feature-based with low task difficulty and is more likely to be object-based with high task difficulty.

Figure 3 shows the capacity estimates using a feature-based unit. Under the low resolution condition, the 95% CIs of the slopes for colors included 1 and their BF_{10} s were greater than 1 for all set sizes (Table 3), indicating that an feature-based unit is appropriate for estimating capacity

for color stimuli. The 95% CIs of the slopes for shape included 1 only for set size 2, which agreed with the BF_{10} s, indicating that VWM for shape might involve different storage units given different task difficulty. Under the high resolution condition, the 95% CIs of all the slopes did not include 1, which also agreed with the BF_{10} s, indicating that a capacity estimates based purely on a feature unit may not be appropriate. However, note that the ranges of most of the 95% CIs were quite close to 1 (e.g., color at set sizes 2 with high resolution) instead of 2, suggesting that the storage unit may be predominantly feature-based. When the capacity was estimated using an object-based unit (Fig. 4), the 95% CIs of most of the slopes did not include 1 except for shapes under the high-resolution condition at set size 6, and its BF_{10} was greater than 3 under this condition (Table 3). This indicates that a capacity estimates based purely on an object unit may not be appropriate under most conditions but with highest task difficulty used in our experiment.

In addition, the results of hit rates comparisons between the conjunctive-feature and single-feature conditions are shown in Table 2. As in Experiment 1, paired-sample t tests (two-tailed, p values were corrected by FDR adjustments) showed that when detecting color changes, the hit rates

Table 3 Line slopes, 95% CIs, r^2 s, and BF_{10} s of the regression analysis

	Slope	95% CI	r^2	BF_{10}		Slope	95% CI	r^2	BF_{10}
Feature-based									
Low resolution					High resolution				
Overall	1.21	[1.14, 1.28]	.59	< .01	Overall	1.40	[1.26, 1.54]	.42	< .01
Color	1.09	[1.00, 1.18]	.50	4.09**	Color	1.28	[1.10, 1.46]	.24	< .01
SS 2	1.04	[1.00, 1.08]	.34	1.36*	SS 2	1.22	[1.10, 1.34]	.30	< .01
SS 4	1.12	[1.00, 1.22]	.53	1.15*	SS 4	1.24	[1.01, 1.47]	.19	.20++
SS 6	1.14	[0.96, 1.34]	.15	1.89*	SS 6	1.38	[1.12, 1.64]	.15	.07
Shape	1.24	[1.12, 1.34]	.27	< .01	Shape	1.44	[1.28, 1.60]	.31	< .01
SS 2	1.01	[0.98, 1.08]	.39	1.51*	SS 2	1.16	[1.06, 1.26]	.76	< .01
SS 4	1.21	[1.06, 1.38]	.17	< .01	SS 4	1.36	[1.10, 1.62]	.17	.03
SS 6	1.36	[1.08, 1.64]	.13	< .01	SS 6	1.78	[1.46, 2.10]	.14	.04
Object-based									
Low resolution					High resolution				
Overall	0.61	[0.57, 0.64]	.59	< .01	Overall	0.70	[0.63, 0.77]	.42	< .01
Color	0.50	[0.50, 0.59]	.50	< .01	Color	0.64	[0.55, 0.73]	.24	< .01
SS 2	0.52	[0.50, 0.54]	.34	< .01	SS 2	0.61	[0.55, 0.67]	.30	< .01
SS 4	0.56	[0.50, 0.61]	.53	< .01	SS 4	0.62	[0.49, 0.74]	.19	.11++
SS 6	0.57	[0.48, 0.67]	.15	< .01	SS 6	0.69	[0.56, 0.82]	.15	.14++
Shape	0.62	[0.56, 0.67]	.27	< .01	Shape	0.72	[0.64, 0.80]	.31	< .01
SS 2	0.51	[0.49, 0.54]	.39	< .01	SS 2	0.58	[0.53, 0.63]	.76	< .01
SS 4	0.61	[0.53, 0.69]	.17	< .01	SS 4	0.68	[0.55, 0.81]	.17	.18++
SS 6	0.68	[0.54, 0.82]	.13	.44+	SS 6	0.89	[0.73, 1.05]	.14	3.57**

Note: **Moderate evidence for H_1 ; *Anecdotal evidence for H_1 ; ++Moderate evidence for H_0 ; +Anecdotal evidence for H_0 ; BF_{10} without superscript indicates strong evidence for H_0

for the conjunctive-feature stimuli were not significantly different from that for the single-feature colors regardless of set size and resolution; when detecting shape changes, the hit rates for the conjunctive-feature stimuli were significantly higher than that for the single-feature shapes except in the low resolution condition with a set size of 2. This was also the case for the average hit rates. As we mentioned in the *Data Analysis* section, a significant increase in the hit rates for the conjunctive-feature stimuli indicated an object-based storage unit in VWM. No other significant difference was found.

Taken together the results of these analysis, we concluded that storage for color tends to be feature-based under low-resolution condition, while storage for shape tends to be feature-based under low-resolution condition at a set size of 2 but to be object-based under high-resolution condition at a set size of 6. For the other situations, storage in VWM is likely to be based on a mixture of feature unit and object unit. The results also suggested a general trend that VWM uses feature as the unit of storage when task difficulty is low but gradually changes to object as the storage unit when task difficulty is high.

Discussion

The present study investigated the effect of task difficulty on the storage unit of visual working memory. Experiment 1 showed that the VWM performance differed for different types of stimuli and with different resolution demands, and the analyses on the hit rates indicated that the storage unit in VWM varied with task difficulty; Experiment 2 showed that the unit of storage in VWM tended to be feature-based with low task difficulty, and to be object-based with high task difficulty. These results challenged fixed-capacity working memory models (Luck & Vogel, 1997; Vogel et al., 2001; Wolters & Raffone, 2008; Zhang & Luck, 2008; Luck & Vogel, 2013), and suggested that the mechanism of VWM is dynamic and dependent on task difficulty.

Our results of Experiment 1 were consistent with previous findings that increasing the resolution demand or featural complexity deteriorates the memory performance (Fougnie et al., 2010; Gao et al., 2013), reduces memory capacity (Alvarez & Cavanagh, 2004; Eng et al., 2005), or decreases the precision of the stored representations (Fougnie et al., 2010; Salmela & Saarinen, 2013). These results seem to contradict a strong object-based storage suggested by the ‘slot’ model, that there are no more than four fixed ‘slots’ in VWM, with each being able to hold one object regardless of its complexity or resolution of representation. Indeed, replications of Luck and Vogel (1997) study lead to mixed results, with success in certain situations and failure in others (Delvenne & Bruyer, 2004; Hardman

& Cowan, 2015; Olson & Jiang, 2002). In addition, evidence shows that averaged representations of objects can be stored in memory to complement object recognition (Dube & Sekuler, 2015), therefore an assumption of independent storage for objects in the slot model is violated. In our Experiment 1, despite the robust findings that memory performance varied inversely with resolution demand, the interpretation of results complicated as one compared the results for single-feature and conjunctive-feature stimuli. In the high-resolution condition, the estimated capacities (assuming an object-based unit) for single-feature colors, shapes, and color-shape conjunctions were approximately the same (Fig. 2a). This is in accordance with the ‘slot’ model. However, in the low-resolution condition, the estimated capacity for colors was greater than for shapes and color-shape conjunctions, but the latter two were not significantly different (Fig. 2a). Moreover, when we split the results for detecting color and shape changes in the conjunctive-feature condition, we found that the hit rates for detecting color changes in the conjunctive-feature stimuli was lower than that in the single-feature colors (Table 2), rejecting an object-based unit account. The results were consistent with that in Experiment 2, even though the comparisons in Experiment 1 and 2 were made on different unit of set size (object as the unit in Experiment 1 and feature as the unit in Experiment 2, see Table 2). The results of both experiments provide solid evidence that the unit of storage in VWM is not fixed.

Therefore, as some researchers suggested, the storage unit in VWM could be a combination of feature-based and object-based representation (Brady & Alvarez, 2011; Delvenne & Bruyer, 2004; Fougnie et al., 2010; Fougnie & Alvarez, 2011; Vergauwe & Cowan, 2015; Wheeler & Treisman, 2002). The key question is under what circumstances the storage of VWM tends to be feature-based or object-based. Our Experiment 2 provides a plausible answer to this question. In this experiment, we manipulated task difficulty by set size, resolution demand, and feature type. Both the single-feature and conjunctive-feature stimuli were tested. Assume that the mechanism involved in temporarily maintaining the visual stimuli is consistent when the same amount of to-be-remembered information was presented and the same task was performed, the estimated memory capacity for single-feature objects and for conjunctive-feature objects should be matched with a proper unit of storage, i.e., the theoretical prediction for the slope of the single-conjunctive fitted line should be 1. We found that when task difficulty is low, e.g., with a small set size, a simple feature and a low resolution demand, the storage unit tended to be feature-based as the 95% CIs of the slopes using a feature-based data set included 1 and the BF_{10} s were greater than 1 under these conditions. As task difficulty increased, the slope

using a feature-based data set deviated farther from 1 while that using an object-based data set gradually approached 1 (Figs. 3 & 4). This variation can be observed as the set size and the resolution demand increase. Specifically, with the largest set size and a high resolution demand, the 95% CI of the slope for detecting shape changes using an object-based data set included 1 and its BF_{10} was greater than 3, indicating an object-based storage. Based on these results, we suggest that the unit of storage in VWM depends on task difficulty—it tends to be feature-based with low task difficulty, and tends to be object-based with high task difficulty. This could be a strategic choice employed by the visual system for optimal deployment of the limited memory resources.

This hypothesis could also explain some seemingly contradictory results found in previous studies. For example, Olson et al. (2002) found the storage unit in VWM could be feature-based in some situations and be object-based in others. In their study, a memory array of color-color conjunctions showed a result pattern consistent of a feature-based storage unit (Experiments 1, 2 & 4, also see Hardman and Cowan, 2015) while a memory array of size-orientation conjunctions showed a result pattern consistent of an object-based storage unit (Experiment 3). Although these results might be explained by separate storage mechanisms for features from different dimensions and same-dimension feature conjunctions (Wheeler & Treisman, 2002), they could be explained by our hypothesis as well. The color-color conjunctions used in their study were similar to the colors in the color categories used in our low-resolution condition, in which task difficulty was low, therefore the storage unit tends to be feature-based, whereas size-orientation conjunctions might indicate high task difficulty since both of the features changed on a continuous scale and required a finer resolution in order to detect a change. This could have resulted in a tendency of object-based storage unit. Our hypothesis is also in line with Vergauwe and Cowan (2015), which suggests that the basic unit of VWM is not fixed. In their study, the authors examined VWM representations of several concurrently held objects and their features by comparing the reaction time of probing with color, shape or both. The stimuli were presented as either conjunctive objects (colored shapes) or single features, and the instructions were varied to explicitly either encourage or discourage the use of binding information. They found feature to be the favored unit for search in VWM in three out of four testing conditions, and that explicitly encouraging binding through instructions made people opt for object-based search only if the probes were presented as binding relevant. They concluded that the basic unit was predominantly feature-based, which was in accordance with our results (but note that the advantage for a feature-based storage in our study may be due to the simple stimuli—color and shape—and thus

generally low task difficulty). Their results also indicate that the unit of representations in VWM includes both features and objects so that, even though different testing situations can result in a stronger emphasis on either the feature dimension or the object dimension, both levels of representation affect memory performance. The authors further suggest that the choice between object-based and feature-based unit seems to be a strategic one, influenced by instructions, and relatively small changes in testing situation can influence the favored unit used for VWM. Compared to their study, our findings further advance our understanding of the representations held in VWM by clarifying the choice of the storage unit in absence of explicit guidance from instructions. That is, the two storage unit of VWM are not mutually exclusive but can be linked through task difficulty—the storage unit tends to be feature-based with low task difficulty and to be object-based with high task difficulty.

One might suspect if other possible explanations could account for our results. It has been suggested that the decrease in memory performance for complex stimuli may be attributed to an increasing difficulty for comparing the memory array and the probe (Awh et al. 2007; Barton, Ester, & Awh, 2009). Complex stimuli (e.g., random polygons) often require a higher resolution of representation because of greater similarity among the memory items as well as between the memory item and the probe, therefore variations in performance reflects a comparison difficulty during retrieval rather than a change in VWM capacity. Indeed, this could account for the worse memory performance found in the high-resolution condition compared to that in the low-resolution condition, since the stimuli and probe used in the high-resolution condition appeared more similar than those in the low-resolution condition. However, it cannot explain the result pattern found in Experiment 2, which shows a divergence in memory performance between the single- and conjunctive-feature stimuli even when the same amount of to-be-stored visual information was presented. Even though the comparison difficulty was different under the different resolution conditions, it remains the same under a fixed resolution condition, i.e., the comparison process and therefore the retrieval stage should be equivalent across single and conjunctive-feature stimuli given the same to-be-stored visual information. The analysis of matching the estimated capacity for the single and conjunctive-feature stimuli allows us to assess the pattern of change in capacity free of the influence of comparison difficulty. The absolute value of memory performance, which might be affected by a comparison process, is not of our primary interest. We focused on the variations in the slope of the fitted line for the two types of stimuli, which indicates that the unit of storage in VWM varies under different testing conditions.

Another concern might be raised about the comparison between the single- and conjunctive-feature displays used in Experiment 2. Indeed, differences can be observed between the two displays: 1) single-feature stimuli occupied twice as many spatial locations as conjunctive-feature stimuli given the same set size of features; 2) task-irrelevant features (gray and square) were presented in single-feature display, but not in conjunctive-feature display. Although there are a few studies demonstrating that these differences do not affect change detection performance under certain conditions (Fougnie et al., 2010), several studies have shown a decrease in memory performance when the same number of features occupy more locations (Delvenne & Bruyer, 2004; Fougnie et al., 2013; Shin & Ma, 2017) and when task-irrelevant features are involved (Shin & Ma 2016, 2017). However, we argue that these differences allowed us to test the unit of storage in VWM, since they were in fact directly related to the two predictions from an object-unit hypothesis we discussed in the *Introduction*. In other words, since the to-be-stored information was equivalent in the single- and conjunctive-displays, whether the irrelevant features or additional locations affected the memory performance would reflect the choice of the storage unit. If an object-based storage unit is in effect, irrelevant features should be automatically encoded (Shin & Ma 2016, 2017) and location information should be registered as it is suggested to be associated with object identity (Treisman & Gelade, 1980). Take Fig. 3 as an example, the more irrelevant features deteriorated the performance and thus affected the capacity, the farther the slope of the corresponding fitted line deviated from one, indicating an object-based storage unit. More importantly, our study showed a change in the results pattern with task difficulty—capacity estimates based on a feature-unit were approximately matched for the single and conjunctive-feature stimuli under low-resolution condition with small set sizes, while they diverged farther as task difficulty increased. In other words, task-irrelevant features (square and gray) and additional locations could be ignored and did not affect the performance with low task difficulty, indicating a feature-based unit; whereas they could not be ignored and thus worsened the performance with high task difficulty.

In addition, cautions should be noted as one interprets the results of Experiment 2. Since the reported VWM capacity in the past literature is 4.4 for color and 2.6 for line drawing (Alvarez & Cavanagh, 2004), one might suspect that in our Experiment 2, the memory performance for the low-resolution colors, shapes, and conjunctions with a set size of 2 could be sufficiently high leaving the range of slopes of the fitted lines truncated. Indeed, the average accuracy was 96% for detecting low-resolution color changes and 92% for detecting low-resolution shape

changes (see Table 1, but no sign of a ceiling effect was found for detecting high-resolution color and shape changes). However, as one scrutinize the data, detecting changes in low-resolution colors demonstrated a tendency of feature-based storage unit regardless of the set size (see 95% CIs and BF_s in Table 3), whereas detecting changes in shapes demonstrated a consistent trend of varying from feature-based unit to object-based unit as the set size increased for both the low- and high-resolution conditions. Therefore, we think that including the results from the set size of 2 in the low-resolution condition could make the case more complete without compromising the validity of the conclusion. Another caveat is that the y-intercept of the fitted line was set to 0, resulting in a greater residual variance and a smaller r^2 than using a liberal y-intercept for linear fitting. However, the latter produces various y-intercepts for different experimental conditions and a great divergence among the slopes, resulting in no basis for comparing the slopes. A y-intercept of 0, which bore a theoretical meaning, was chosen to provide a uniform start point for fitting and comparing the slopes. Additionally, individual differences might also contribute to the large residuals. We could evaluate task difficulty in general given the observers' average performance, but it is possible that the degree of difficulty for a certain task varies for individual observers. In other words, the same task can be difficult for one while be easy for another. As a result, the criteria for whether to adopt a more feature- or more object-based storage in a certain task could be different for individual observers and therefore the goodness of fits was less satisfying in our experiment.

Finally, besides the aforementioned behavioral studies that provide us insights into the underlying mechanisms of VWM, neurophysiological studies also provide evidence for the neural substrates of visual memory processes. It is known that an event-related potential component, contralateral delay-related activity (CDA; Vogel & Machizawa, 2004), reflects the encoding and maintenance of items in visual memory. The amplitude of CDA increases with the number of objects being held in the memory, and plateaus at a set size that predicts one's VWM capacity limits (Vogel & Machizawa, 2004). In a recent study, Brady, Störmer, and Alvarez (2016) found CDA evidence for better performance and more active storage capacity for real-world (more complex) objects in WM, suggesting that capacity in WM is not fixed but dependent upon our existing knowledge. In addition, mixed results were found on whether the resolution of representation affects the capacity of VWM at the maintenance phase or at the retrieval phase (Awh et al. 2007; Gao et al. 2013; Machizawa, Goh, & Driver, 2012). Gao and colleagues found that there was no difference in the CDA amplitudes under different resolution conditions, indicating that the effect of resolution does not

occur at the maintenance phase. However, a few studies showed the opposite (Machizawa et al. 2012; Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010). Machizawa found higher CDA amplitudes as the required precision increased, but only with a small number of retained items. Functional magnetic resonance imaging (fMRI) data show that posterior parietal cortex is tightly correlated with the capacity of VWM, where the activity increases with set size (Todd & Marois, 2004). Xu (2006) found that the inferior intraparietal sulcus (IPS) maintains a fixed number of object representations at different spatial locations regardless of object complexity, and the superior IPS and the lateral occipital complex encode and maintain a variable subset of the attended objects, representing fewer objects as their complexity increases. However, the underlying mechanisms of VWM are not fully understood. Further neurophysiological research may shed lights on the storage unit of VWM with regard to different levels of task difficulty.

Acknowledgments This work has been supported by the National Natural Science Foundation of China (31500919).

Open Practices Statement The data and materials for all experiments are available upon request and none of the experiments was preregistered.

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(2), 106–111.
- Awh, E., Barton, B., & Vogel, E. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622–628.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829.
- Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1359.
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 7–7.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321, 851–854.
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49(6), 1622–1631.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392.
- Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, 113(27), 7459–7464. <http://www.pnas.org/content/113/27/7459.abstract>. <https://doi.org/10.1073/pnas.1520027113>
- Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, 24(01), 154–176.
- Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition*, 11(1), 1–27.
- Dube, C., & Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *Journal of Vision*, 15(4), 1–13.
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, 12(6), 1127–1133.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, 11(12), 3–3.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12), 27–27.
- Fougnie, D., Cormiea, S. M., & Alvarez, G. A. (2013). Object-based benefits without object-based representations. *Journal of Experimental Psychology: General*, 142(3), 621–626.
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, 17(5), 673–679.
- Gao, T., Gao, Z., Li, J., Sun, Z., & Shen, M. (2011). The perceptual root of object-based storage: An interactive model of perception and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 37(37), 1803–1823.
- Gao, Z., Ding, X., Yang, T., Liang, J., & Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PloS One*, 8(2), e57913.
- Gao, Z., Li, J., Liang, J., Chen, H., Yin, J., & Shen, M. (2009). Storing fine detailed information in visual working memory—evidence from event-related potentials. *Journal of Vision*, 9(17), 1–12.
- Gegenfurtner, K. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, 4(7), 563–572.
- Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 325.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, 114(3), 599.
- Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Attention, Perception, and Psychophysics*, 70(8), 1581–1591.
- Lee, D., & Chun, M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, 63(2), 253–257.
- Luck, S. J., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Luck, S. J., & Vogel, E. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400.
- Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, 22(3), 496–512.
- Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, 49, 1632–1639.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347.
- Machizawa, M. G., Goh, C. C., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, 23(6), 554–559.

- Magnussen, S., & Greenlee, M. (1999). The psychophysics of perceptual memory. *Psychological Research*, 62(2), 81–92.
- McCollough, A. W., Machizawa, M. G., & Vogel, E. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex*, 43(1), 77–94.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of “strong object” hypothesis. *Perception & Psychophysics*, 64(7), 1055–1067.
- Parra, M. A., Sala, S. D., Logie, R. H., & Morcom, A. (2014). Neural correlates of shape–color binding in visual working memory. *Neuropsychologia*, 52, 27–36.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44(4), 369–378.
- Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, 139(6), 1173.
- Qian, J., Zhang, K., Lei, Q., Han, Y., & Li, W. (2019). Task-dependent effects of voluntary space-based and involuntary feature-based attention on visual working memory. *Psychological Research*, 1–16. (Online publication).
- Salmela, V. R., & Saarinen, J. (2013). Detection of small orientation changes and the precision of visual working memory. *Vision Research*, 76, 17–24.
- Shen, M., Yu, W., Xu, X., & Gao, Z. (2013). Building blocks of visual working memory: Objects or Boolean maps? *Journal of Cognitive Neuroscience*, 25(5), 743–753.
- Shin, H., & Ma, W. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5), 10–10.
- Shin, H., & Ma, W. (2017). Visual short-term memory for oriented, colored objects. *Journal of Vision*, 17(9), 12–12.
- Thornton, T. L., & Gilden, D. (2007). Parallel and serial processes in visual search. *Psychological Review*, 114(1), 71.
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984), 751–754.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neurosciences*, 27(5), 853–865.
- Vergauwe, E., & Cowan, N. (2015). Working memory units are all in your head: Factors that influence whether features or objects are the favored units. *Journal of Experimental Psychology: Learn Memory and Cognition*, 41(5), 1404–1416.
- Vogel, E. K., & Machizawa, M. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751.
- Vogel, E. K., Woodman, G. F., & Luck, S. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92.
- Wang, B., Cao, X., Theeuwes, J., Olivers, C. N., & Wang, Z. (2016). Location-based effects underlie feature conjunction benefits in visual working memory. *Journal of vision*, 16(11), 12–12.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48.
- Wilson, K. E., Adamo, M., Barense, M. D., & Ferber, S. (2012). To bind or not to bind: Addressing the question of object representation in visual short-term memory. *Journal of Vision*, 12(8), 14–14.
- Wolters, G., & Raffone, A. (2008). Coherence and recurrency: Maintenance, control and integration in working memory. *Cognitive Processing*, 9(1), 1–17.
- Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, 64(8), 1260–1280.
- Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 458.
- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, 68(5), 815–828.
- Ye, C., Zhang, L., Liu, T., Li, H., & Liu, Q. (2014). Visual working memory capacity for color is independent of representation resolution. *PloS One*, 9(3), e91681.
- Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.