

THE CAPACITY AND PRECISION OF VISUAL WORKING MEMORY FOR OBJECTS AND ENSEMBLES

YU.A. MARKOV^a, N.A. TIURINA^a, YU.M. STAKINA^a, I.S. UTOCHKIN^a

^a *National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow, 101000, Russian Federation*

Abstract

Previous research has documented the limited capacity of visual working memory (VWM) for color objects set at 3–5 items. Another line of research has shown that multiple objects can be stored in a compressed form of ensemble. However, existing data is more likely to testify that VWM can store no more than two such compressed units. But the nature of this discrepancy can be methodological: VWM for ensembles was never tested using methods that are applied in the research of VWM for objects. Here we have tested the capacity and precision of VWM for objects and ensembles using two standard methods — change detection and continuous report with a mixture model. We found that VWM for both types of units showed the similar capacity and precision when critical psychophysical parameters, such as foveal density and area are controlled. We also showed that this quantitative similarity between objects and ensembles is provided by a mechanism that represents each ensemble as a holistic VWM chunk as efficiently as it represents any single object.

Keywords: visual working memory, object perception, ensemble perception.

Working memory is often referred to as a system that actively maintains and operates information necessary for current goals and tasks (Baddeley, 1986; Baddeley & Hitch, 1974). One of the most important attributes of working memory is its limited capacity, the maximum number of separate representations that are concurrently maintained in the system. Across numerous tasks, modalities, and conditions, the average capacity is shown to be about

four units (Cowan, 2001), however, it also shows some individual differences between people (Luck & Vogel, 2013).

Methods for studying capacity and precision of VWM for objects

Within a domain of visual working memory (VWM), capacity limits are also established. In their seminal work, Luck and Vogel (1997) claimed that three-four individual items can be

stored in memory, and these items are individual objects. They used a version of a change detection paradigm (Pashler, 1988; Phillips, 1974). The typical change detection task consists of a briefly presented sample containing a variable number of objects, a blank interval when the sample should be stored in VWM, and a test display that can be exactly the same as the sample or having one item changed. The observer should determine whether the change is present or absent. Using detection accuracy as a function of the number of objects, an actual VWM capacity can be estimated (Cowan, 2001). In their change detection study, Luck and Vogel (1997) found that people were equally good at detecting a single change in a set of objects varying in only one dimension (color) and at detecting a change among the same number of objects varying in four dimensions (color, orientation, size, and the presence/absence of a gap). However, later research questioned this conclusion showing that the capacity to detect change strongly depends on the heterogeneity and complexity of material to be stored (Alvarez & Cavanagh, 2004; Olson & Jiang, 2002; Wheeler & Treisman, 2002).

In attempt to address the controversial change detection data, Wilken and Ma (2004) suggested using a continuous report task as a strong addition to the discrete response system used for change detection. In their paradigm, participants memorized a sample display and, after retention, had to adjust the color of a single probed item from that display to match the original color of the sample item in the same location. The distribution of errors (response deviations from the true color) is then

analyzed, and its standard deviation is accepted as a measure of VWM precision. Combining this method with change detection, Wilken and Ma (2004) concluded that VWM capacity is limited by the noise increasing with additional items and reducing the precision of each individual item. However, Zhang and Luck (2008) suggested a different approach to the analysis of error distribution based on mixture modeling. Armed with this method, Zhang and Luck (2008) separated two types of errors: random guessing (reporting values that are not in memory, which produces a uniformly distributed component of the model) and an imprecise report of an item that is in memory (which produces clustering errors around the true value in the form of a normal distribution). Calculating the standard deviation (SD) for a normally distributed component of the model seems to be a more correct way to estimate the precision of an item that is really stored in VWM. Also, the total area of the random guess distribution can be used to determine how many items are in fact in memory that is its exact capacity.

VWM for ensembles vs. objects

While capacity for individual objects is severely limited (Cowan, 2001; Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 1997), there seem to be strategies that the visual system uses to bypass these limitations. One such strategy can rely on natural regularities of the stimulus to form compressed representations of multiple objects. It is shown in numerous experiments that observers can successfully extract such compressed representations in a form of

ensemble summary statistics across various sensory (Alvarez & Oliva, 2009; Ariely, 2001; Bauer, 2009; Chong & Treisman, 2003; Dakin & Watt, 1997; Watamaniuk & Duchon, 1992) and even high-level perceptual (Haberman & Whitney, 2007, 2009; Yamanashi Leib, Kosovicheva, & Whitney, 2016) dimensions. The phenomenon of ensemble summary statistics consists in reasonably rapid (Chong & Treisman, 2003; Robitaille & Harris, 2011; Whiting & Oriet, 2011) and precise (Alvarez, 2011) judgment of the average parameter of multiple objects. The idea of VWM compression using ensemble summaries implies that observers do not memorize the full number of objects with great precision but can retrieve some information about any object using the general summary (Brady & Alvarez, 2011; Corbett, 2017). The quality of retrieval would be inevitably worse than when each single object is encoded. But still, the estimates would be better than mere random guessing even when the number of objects overcomes the known limits of VWM (Corbett, 2017).

When individual items become organized into an ensemble, they are likely to form a single unit for attention and working memory (Corbett, 2017; Im & Chong, 2014; Im, Park, & Chong, 2015), which means that the quality of ensemble encoding marginally depends on the number of individuals within (Ariely, 2001, 2008; Attarha & Moore, 2015; Attarha, Moore, & Vecera, 2014; Chong, Joo, Emmanouil, & Treisman, 2008; Robitaille & Harris, 2011; Utochkin & Tiurina, 2014; but see Marchant, Simons, & De Fockert, 2013; Maule & Franklin, 2016; Myczek & Simons, 2008; Simons & Myczek,

2008). However, the number of such ensemble units can be limited.

Some studies addressed the issue of VWM capacity for multiple objects organized in ensemble fashion. Chong & Treisman (2005) were the first to show that ensemble features (mean sizes) can be extracted at one time from at least two spatially overlapping sets. However, they did not test more than two such sets. Im and Chong (2014) moved further and tested an ability to estimate the mean sizes of up to five ensembles. They found that the accuracy steadily declines starting with three sets. This result shows that the capacity limit is very low – probably no more than two. Attarha and Moore (2015; Attarha, et al., 2014) presented four ensembles either simultaneously or sequentially (two at a time) and found that the sequential method (when VWM is loaded by only two objects at one time) provides better performance, which is also consistent with the limited VWM capacity for ensembles of about two units. This estimate is supported by the data from experiments on approximate estimation of numerosity – another statistical summary of multiple objects. Halberda, Sires, and Feigenson (2006) reported that their participants could estimate an approximate number of dots in two color subsets without loss in precision, even when they did not know in advance which subsets they would be asked about.

Further experiments showed that the limit of ensemble memory probably arises not from limits in the “processor” computing ensemble properties, such as the mean feature or numerosity. In the most reported studies (except for Attarha & Moore, 2015, and Attarha et

al., 2014), ensembles were presented as spatially overlapping subsets, with objects from one subset intermingling with objects from other subsets. Such a way of presentation makes ensembles different from single items and spatially grouped sets because they have no clear spatial boundaries that would provide the “objecthood” of each subset (Trick & Pylyshyn, 1993). Using such stimuli, Poltoratski and Xu (2013) replicated the no-more-than-two finding from Halberda et al. (2006) and then, using the partial report, showed that this limit reflects a failure to encode more than two colors (subset-defining features) to VWM rather than computational limits of number estimation. Earlier, Watson, Maylor, & Bruce (2005) came to a similar conclusion and similar capacity estimate when asked participants to report the number of color subsets and measured their reaction time.

Our study

From the previous section we see that evidence accumulated from different paradigms converge to provide a conclusion that representing ensembles in VWM is capacity-limited, and that this limit is very severe – probably around two ensembles at one time. Here, we see a discrepancy between the estimated capacity for ensembles and individual objects. Given that in most of the experiments ensembles were defined by colors, ensemble capacity seems substantially lower than object capacity that is set closer to three-four (Luck & Vogel, 1997) or even five (Alvarez & Cavanagh, 2004) colors.

In our study, we addressed the discrepancy between the estimated capac-

ities of VWM for ensembles and individual objects. We see a very important problem in that VWM for ensembles was never tested by the standard methods typically used in contemporary studies of VWM for objects, such as change detection and continuous report. The tasks used in the ensemble studies (see previous section) are different in terms of their demands – report the average (Attarha & Moore, 2015; Attarha et al., 2014; Chong & Treisman, 2015; Im & Chong, 2014) or the number (Halberda et al., 2006; Poltoratski & Xu, 2013; Watson et al., 2005), which probably involves more complex operations than just retention and retrieval of multiple colors. Partial color reporting used by Poltoratski and Xu (2013) is closer to standard VWM tests but also somewhat more difficult: while both change detection and continuous report keep the spatial reference of a tested item, Poltoratski and Xu’s (2013) method did not, which could complicate retrieval. Moreover, the precision of ensemble encoding was never measured, since the continuous report paradigm has never been applied to ensembles in a way as it is applied to objects.

Our aim, therefore, is to test VWM for both objects and ensembles using exactly the same standard methods and directly compare the corresponding parameters. In Experiment 1, we tested VWM for displays consisting of one to five individual objects, each having a unique color, as compared to displays consisting of one to five overlapping ensembles, each in turn including several objects of a common color. In Experiment 2, we repeated Experiment 1 controlling for the total area of objects and ensembles. In Experiment 3, we

also repeated Experiment 1, but this time we controlled display density and eccentricity from the fovea. Finally, in Experiment 4, we tested whether VWM parameters for ensembles can be accounted for by a sampling strategy that implies selective encoding of a few individual representatives instead of ensembles.

Experiment 1

Method

Participants

Twelve psychology students of the Higher School of Economics (11 female; age: $M = 19.45$ years, $SD = 0.52$) took part in the experiment for extra course credits. All participants reported having normal color vision, normal or corrected to normal visual acuity, and no neurological problems. Before the beginning of the experiment, they signed an informed consent form. One participant's data were excluded from analysis because she showed nearly 100% guess rate in the change detection task.

Apparatus and stimuli

Stimulation was developed and presented using PsychoPy (Pierce, 2007) for Linux. Stimuli were presented on a standard VGA monitor in a refresh frequency of 75 Hz with 1024×768-pixel spatial resolution. Stimuli were presented against a homogeneous gray field. Participants sat at approximately 47 cm from the monitor. From that distance, the screen subtended at approximately 44.7×34.2 degrees of visual angle.

Sample displays. Sets of color circles were generated within a square region

subtending 24.3 degrees and having a center at a fixation point. The diameter of each circle randomly varied between 0.4 and 0.7 degrees. The circles were randomly located within this square region with the only restriction being that they could not overlap. For testing memory of individual objects, one to five circles could be presented, each having a unique color.

For testing memory for ensembles, one to five sets of circles could be presented. Each set consisted of six to eight circles sharing a common color. As all circles were randomly located in the space, the color sets overlapped, that is, circles of one color were interspersed with the circles of different colors (except for displays where only one subset was present).

We used an HSV (hue-saturation-value) palette for coloring the circles in our displays. Both saturation and value were set at their maximum 1, providing that only the hue was variable. We used the following algorithm for assigning hues to objects or sets. For each display, a random hue was first picked from the HSV color wheel and assigned to one of the sets. For the rest of the objects or sets (if more than one is presented), hues rotated by $n \cdot 60 \pm 15$ degrees away from the initial one (where n is an integer multiplier from 2 to 5) could be assigned. This algorithm provided distinctiveness between any two colors no less than at least 30 degrees along the HSV color wheel.

Test displays. In the change detection task, test displays could either be the exact copies of samples or having one object or one set changed in color. In the continuous report task, test displays originally included the outlines of the sample circles without color. One

probed object or set had thicker outlines. This outline layout was surrounded by a hue color wheel (internal and external diameters were 27.3 and 31.4 degrees, respectively) used to select the hues for adjusting the color of the probed object or set.

Procedure

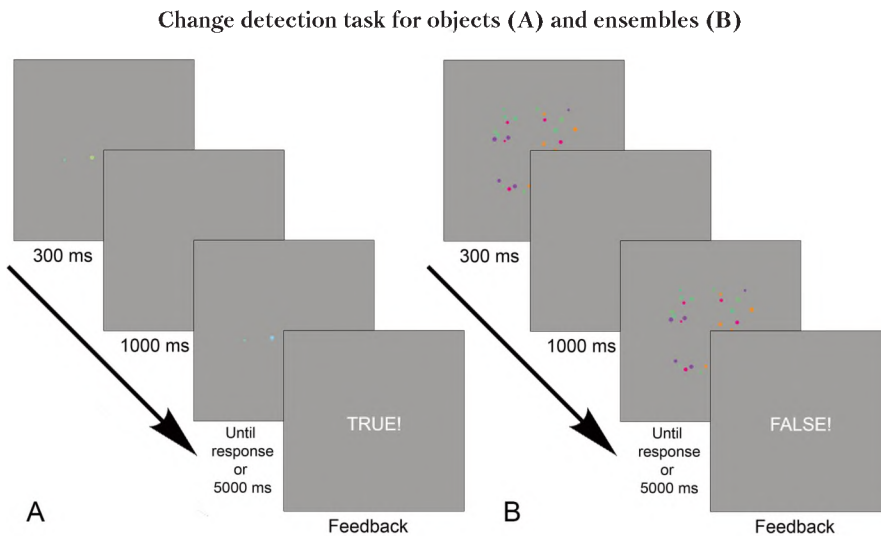
During the experiment, each participant underwent two types of tasks, each having “object” and “ensemble” versions: (1) change detection for objects, (2) change detection for ensembles, (3) continuous report for objects, and (4) continuous report for ensembles. The order of the tasks varied across participants. Each task started with a short practice block.

Change detection. In the change detection task (Figure 1), participants were instructed to memorize the colors of objects or sets presented in sample displays and report whether one of the colors had changed in the test display.

Each trial started with a presentation of a sample display for 300 ms. A 1,000-ms blank interval then followed requiring the participants to retain the sample in memory. After the blank interval, a test display appeared until response or 5,000 ms, whichever occurred earlier. For response, a standard computer keyboard was used. Participants had to press <l> button if they saw a change between the sample and test displays, or <s> if they did not see any change. There was a 0.5 probability of change presence. Feedback was provided after the response whether it had been correct or incorrect. The feedback stayed on the screen until the participant pressed a space bar on the keyboard to start the next trial.

Continuous report. In the continuous report task (Figure 2), participants also had to memorize the colors of objects or sets in a sample presented for 300 ms. After a 1,000-ms blank interval, a test display, as described above, appeared. Clicking on the color wheel with a

Figure 1



computer mouse the participants had to pick the hue corresponding to the hue of the probed object or set. The first click on the color wheel caused the outline probed object or set to take the picked hue. The participants could then correct their response by another click or dragging the mouse. To confirm their final response, the participants had to press the space bar. Feedback was then presented showing how close participant's response was to the correct answer. The feedback was provided by showing two color circles: the color of the left circle corresponded to the true color of a sample, and the color of the right circle corresponded to the participant's response; in addition, a black arrow indicated the true sample color on the color wheel and a white arrow indicated the participant's response, so participants could see the angular distance between their responses and correct responses. The feedback stayed on the screen until the participant pressed a space bar on the keyboard to start the next trial.

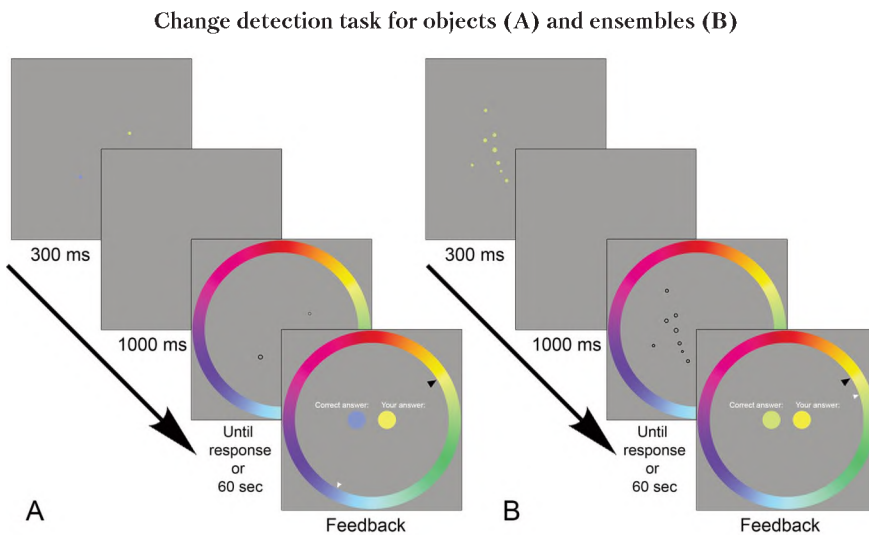
Design and analysis

Two factors were manipulated in this experiment. The first one was Unit Type (two conditions: objects vs. ensembles). The second one was Set Size, the number of objects or ensembles on the screen (five conditions: one to five). We used 50 trials per cell of the factorial design. Therefore, in each of two tasks, every participant took $2 \times 5 \times 50 = 500$ trials.

In the change detection task, we measured the capacity of VWM in each of the set sizes and unit types using Cowan's K formula (Cowan, 2001): $K = (p(\text{Hit}) - p(\text{FA})) * N$, where K is an average estimate of the number of units stored in memory in a given condition, $p(\text{Hit})$ is the probability of "hits" (correct detection when the change is present), $p(\text{FA})$ is the probability of "false alarms" (false detection when the change is absent), N is the set size.

For the continuous report task, errors were calculated in each trial. The error is an angular difference between

Figure 2



the participant's response and the true sample hue of a probed object or ensemble. We then analyzed the distribution of errors in each condition using the mixture model as described by Zhang and Luck (2008). The model separates two basic distributional components: a component of the von Mises distribution (which reflects responses based on reporting colors that are in memory) and a component of the uniform distribution (which corresponds to random guess reflecting the absence of a tested item in memory). We ran mixture models using MemToolBox for Matlab (Suchow, Brady, Fournie, & Alvarez, 2013). From the mixture models, we derived two important parameters. The standard deviation (SD) of the von Mises component was the measure of VWM precision for colors that are in memory. The area of the uniform component reflecting the overall probability of random guess P_{guess} was used to calculate the VWM capacity: $C = (1 - P_{guess}) * N$, where C is an average estimate of the number of units stored in memory, $(1 - P_{guess})$ is the probability that a tested object or ensemble is in memory, N is the set size.

Results

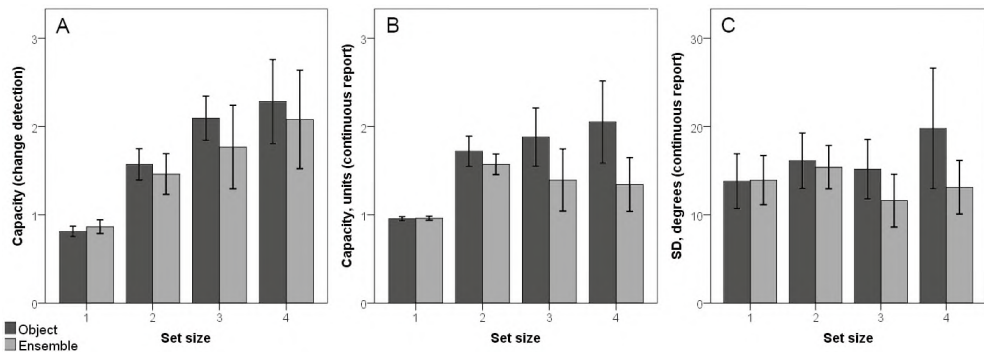
In many participants, the mixture model failed to converge for the set size of five items, which shows that this condition could probably be too difficult. We therefore decided not to include this set size into analysis in this and the following experiments.

The change detection task (Figure 3A) yielded no significant difference between object and ensemble capacities ($F(1, 10) = 2.217, p = 0.167, \eta_p^2 = 0.181$). The effect of the set size was significant ($F(3, 30) = 43.386, p < .001, \eta_p^2 = 0.813$). There were significant differences between the set size = 1, and all the rest of the conditions (p 's < 0.001, Bonferroni corrected). The difference was also significant between the set size = 2 and the set sizes = 3 and 4 ($p < 0.001, p = 0.003$, Bonferroni corrected). The set size \times unit type interaction was not significant ($F(3, 30) = 1.106, p = 0.362, \eta_p^2 = 0.100$).

The continuous report task (Figure 3B-C) showed a significant difference between object and ensemble capacities

Figure 3

Capacity and SD data from Experiment 1. A – change detection task; B, C – continuous report task. Error bars denote 95% CI



and precisions ($F(1, 10) = 23.520, p < 0.001, \eta_p^2 = 0.702$; $F(1, 10) = 5.805, p = 0.037, \eta_p^2 = 0.367$, respectively). The effect of set size for capacity again was significant ($F(3, 30) = 18.751, p < 0.001, \eta_p^2 = 0.652$). There were significant differences between set size = 1 and all of the rest set sizes (p 's < 0.001 , Bonferroni corrected). The set size \times unit type interaction for capacity was significant ($F(3, 30) = 8.183, p < 0.001, \eta_p^2 = 0.450$). There were significant differences between object capacity and ensemble capacity for set size = 3 and 4 ($p = 0.005, p = 0.003$). For precision, there were no effects of set size ($F(3, 30) = 1.728, p = 0.182, \eta_p^2 = .147$) and of set size \times unit type ($F(3, 30) = 2.153, p = 0.114, \eta_p^2 = 0.177$).

Experiment 2

As ensembles were more numerous than individual objects in Experiment 1, they were distributed more densely on the screen providing more chance for any region to be filled with some items. This could lead to a higher probability of at least a few items falling into the fovea, which is important for precise encoding of color. This could explain why we observed a higher capacity in the ensemble condition of Experiment 1. To address this issue, in Experiment 2 we equated the foveal density of objects and ensembles.

Method

Participants

Twelve psychology students of the Higher School of Economics (10 female; age: $M = 19.58$ years, $SD = 0.79$) took part in the experiment for extra

course credits. All participants reported having normal color vision, normal or corrected to normal visual acuity, and no neurological problems. Before the beginning of the experiment, they signed an informed consent form. Three participants' data were excluded from analysis because they showed a nearly 100% guess rate in set size = 3 and 4.

Apparatus and stimuli

Apparatus and stimuli were the same as in Experiment 1, but with one important exception. In the individual object condition, location coordinates for the circles were generated within a narrower region around fixation (4.59 degrees). This provided approximately the same foveal density as in the ensemble condition, for which coordinates were generated within the same area as in Experiment 1.

Procedure, design, and analysis were exactly the same as in Experiment 1.

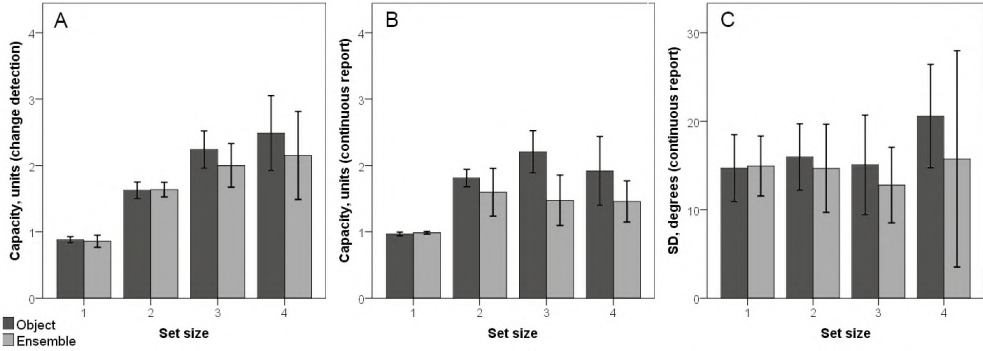
Results

Change detection (Figure 4A) again showed no difference between object and ensemble capacities ($F(1, 8) = 4.973, p = 0.056, \eta_p^2 = 0.383$). The effect of set size was significant ($F(3, 24) = 32.836, p < 0.001, \eta_p^2 = .804$). There were significant differences between the set size = 1 and all of the rest conditions (p 's < 0.001 , Bonferroni corrected) and also between conditions with the set size = 2, and conditions with the set size = 3 and 4 ($p = 0.032, p = 0.001$, Bonferroni corrected). The set size \times unit type interaction was not significant ($F(3, 24) = 1.244, p = 0.316, \eta_p^2 = 0.135$).

The continuous report task (Figure 4B) showed a significant difference

Figure 4

Capacity and SD data from Experiment 2. A – change detection task; B, C – continuous report task. Error bars denote 95% CI



between object and ensemble capacities ($F(1, 8) = 20.026, p = 0.002, \eta_p^2 = 0.715$). The effect of the set size was significant ($F(3, 24) = 17.182, p < 0.001, \eta_p^2 = 0.682$). There were significant differences between condition with the set size = 1, and all the rest conditions (p 's < 0.001, Bonferroni corrected). The set size \times unit type interaction was significant ($F(3, 24) = 6.283, p = 0.003, \eta_p^2 = 0.440$). There was a significant difference between object and ensemble capacities for the set size = 3 ($p < 0.001$). For precision, no significant effects of the set size ($F(3, 24) = 0.842, p = 0.484, \eta_p^2 = 0.095$), unit type ($F(1, 8) = 1.821, p = 0.214, \eta_p^2 = 0.185$), or set size \times unit type ($F(3, 24) = 0.654, p = 0.588, \eta_p^2 = 0.076$) were found (Figure 4C).

Experiment 3

In this experiment we addressed another potential psychophysical confound that could arise between individual objects and ensembles in Experiment 1. As ensembles were more numerous, their total area was larger than the area of the objects. Here we equat-

ed the areas between objects and ensembles.

Method

Participants

Twelve psychology students of the Higher School of Economics (10 female; age: $M = 19.41$ years, $SD = 0.68$) took part in the experiment for extra course credits. All participants reported having normal color vision, normal or corrected to normal visual acuity, and no neurological problems. Before the beginning of the experiment, they signed an informed consent form.

Apparatus and stimuli

Apparatus and stimuli were the same as in Experiment 1, but with another important exception. In the individual object condition, we multiplied the diameters of circles originally used in Experiment 1 by 7, so that the new diameters ranged between 0.978 and 1.738 degrees. This led the average area of the individual circles to become

seven time as big as the average area of the circles in Experiment 1. In the ensemble condition, areas remained the same as in Experiment 1. As the average number of circles in each ensemble was seven, their total area was approximately equal to the area of the magnified individual objects.

Procedure, design, and analysis were exactly the same as in Experiment 1.

Results and discussion

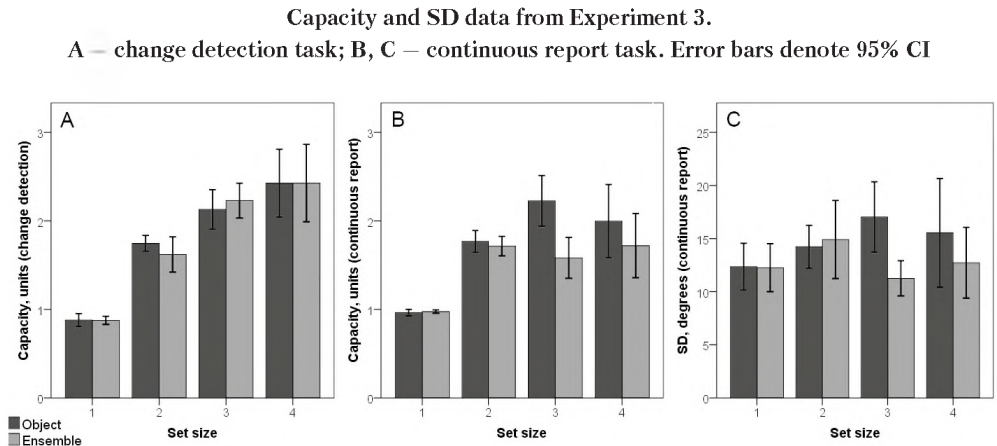
As in previous experiments, change detection (Figure 5A) yielded no significant difference between object and ensemble capacities ($F(1, 11) = .015, p = .906, \eta^2_p = .001$). The effect of set size was significant ($F(3, 33) = 74.549, p < .001, \eta^2_p = .871$). There were significant differences between the set size = 1 and all of the rest conditions (p 's < 0.001 , Bonferroni corrected). And also between conditions with set size = 2, and conditions with the set size = 3 and 4 ($p < 0.001, p < 0.001$, Bonferroni corrected). The set size \times unit type interaction was not significant ($F(3, 33) = 0.608, p = 0.615, \eta^2_p = 0.052$).

The continuous report task (Figure 5B) showed a significant difference between object and ensemble capacities ($F(1, 11) = 8.257, p = 0.015, \eta^2_p = 0.492$). The effect of the set size was significant ($F(3, 33) = 38.700, p < 0.001, \eta^2_p = 0.779$). There were significant differences between condition with the set size = 1, and all the rest conditions (p 's < 0.001 , Bonferroni corrected). The set size \times unit type interaction was significant ($F(3, 33) = 6.036, p = 0.002, \eta^2_p = 0.354$). There was a significant difference between object and ensemble capacities for the set size = 3 ($p < 0.001$). For precision, no significant effects of set size ($F(3, 33) = 2.620, p = 0.067, \eta^2_p = 0.192$), unit type ($F(1, 11) = 9.946, p = 0.009, \eta^2_p = 0.475$), or set size \times unit type ($F(3, 33) = 1.968, p = 0.138, \eta^2_p = 0.152$) were found (Figure 5C).

Experiment 4

In Experiments 2 and 3 we controlled for physical stimulus factors, such as foveal density and area, and found basically the same VWM parameters for both objects and ensembles.

Figure 5



This result might lead us to conclude that VWM treats ensembles as exactly the same units as individual objects. However, does this mean that an ensemble is indeed encoded as a single unit in its entirety? The answer is not obvious. It is possible that only a limited sample of individuals is picked out from the entire number of multiple objects and encoded into VWM, while other ensemble members are not encoded at all. In Experiment 4, we externally controlled the allocation of attention towards individual representatives of ensembles to test whether they bias VWM in favor of the corresponding ensembles. To manipulate the allocation of attention, we used a modification of the abrupt onset paradigm when a single item (Yantis & Jonides, 1984) or a group of items (Jiang, Chun, & Marks, 2002) captures involuntary attention by asynchronous presentation with the rest of a set. Our main idea was to compare VWM parameters for ensembles whose individual representatives are attended with ensembles whose representatives are unattended. In Experiment 4, we tested it only on the continuous report task for ensembles.

Method

Participants

Eighteen psychology students of the Higher School of Economics (16 female; age: $M = 19.44$ years, $SD = 0.78$) took part in the experiment for extra course credits. All participants reported having normal color vision, normal or corrected to normal visual acuity, and no neurological problems. Before the beginning of the experiment, they

signed an informed consent form. Six participants' data were excluded from analysis because they showed nearly a 100% guess rate in all conditions.

Apparatus and stimuli

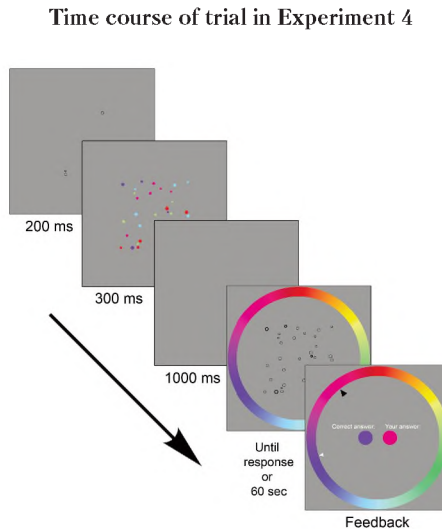
Apparatus was the same as in Experiment 1. We used only one subset of stimuli from Experiment 1, namely, those used for the continuous report task and the ensemble condition. The set size was fixed at five ensembles.

Procedure, design, and analysis

In general, the procedure was the same as described in the Continuous report section of Experiment 1. However, there was an important addition. 200 ms before the entire sample presentation, a subset of one to four objects from that sample appeared and stayed until the sample offset (Figure 6). Each object in a subset had a unique color. After the 1,000-ms retention interval, the observers had to set the color of a probed ensemble. Critically, the probed ensemble could be either one that had a representative in the precued subset, or one that had no such representative.

In this experiment, we manipulated two factors. The first was Sample Size, the number of the cued circles (1, 2, 3, or 4). The second factor was Representativeness: a representative sample always included one item from a subsequently probed ensemble; a non-representative sample had no members of the probed ensemble. Data were analyzed using the mixture model (Zhang & Luck, 2008). As in the previous experiments, capacity (C) and precision (SD) were our target parameters.

Figure 6



Results

The effect of set size, representativeness, and set size \times representativeness interaction were not significant for capacity and precision (capacity: $F(3, 33) = 0.913, p = 0.445, \eta^2_p = 0.077$; $F(1, 11) = 4.475, p = 0.058, \eta^2_p = 0.289$; $F(3, 33) = 1.237, p = 0.312, \eta^2_p = 0.101$; precision: $F(3, 33) = 1.160, p = 0.340, \eta^2_p = 0.095$; $F(1, 11) = .033, p = 0.858, \eta^2_p = 0.003$; $F(3, 33) = .821, p = 0.492, \eta^2_p = 0.069$; Figure 7).

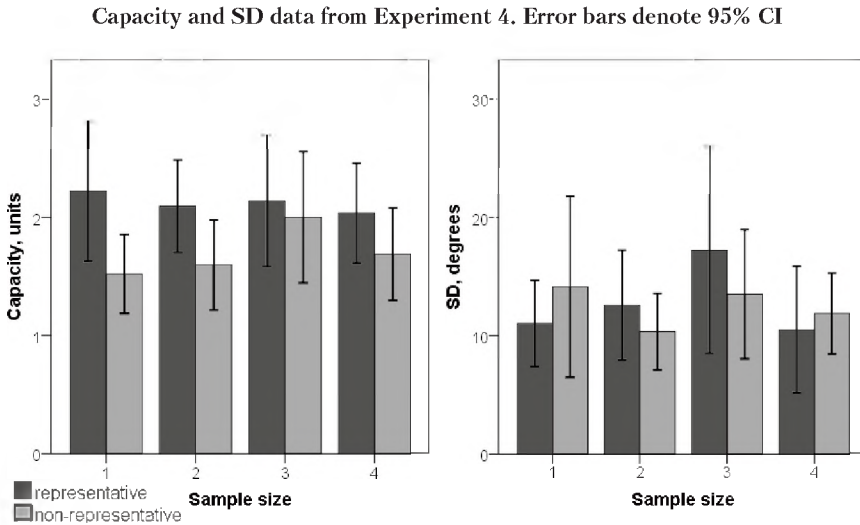
General Discussion

In the series of experiments reported here we investigated VWM for two types of perceptual units – individual objects and spatially overlapping ensembles. The study was inspired by a discrepancy in the literature about the capacities of VWM for individual objects (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997, 2013; etc.) and ensembles (Attarha & Moore, 2015;

Attarha et al., 2014; Halberda et al., 2006; Im & Chong, 2014; Poltoratski & Xu, 2013; Watson et al., 2005). One possible concern about this discrepancy is that it could be caused by differences in methodologies used for measuring the capacity in the studies of object VWM and ensemble VWM. In our study, we compared VWM capacities for objects and ensembles using exactly the same methods. These methods are recognized as standard in the field of object VWM – change detection (Luck & Vogel, 1997) and continuous report (Wilken & Ma, 2004) with the mixture model (Zhang & Luck, 2008).

In general, our findings support an idea that VWM has approximately the same capacities for individual objects and ensembles. Moreover, ensembles tend to be encoded with even higher precision than individual objects with exactly the same properties as ensemble constituents (Experiment 1). This demonstrates a sort of a redundancy

Figure 7



gain caused by object numerosity (Utochkin, 2016). However, when objects and ensembles are equated in their low-level properties, such as foveal density or area, precision becomes the same for both objects and ensembles (Experiments 2 and 3).

Low VWM capacity for individual objects

One result of our study seems to be a bit challenging given the data described in the literature. Experiments with change detection as the basic paradigm to measure VWM capacity for objects usually show greater limit numbers – at least three-four (Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997) or even almost five (Alvarez & Cavanagh, 2005) items when observers are asked about the color. Using the same method, we came to much lower estimates: in our Experiments 1-3, change capacity did not exceed two items.

One possible explanation for this rather big difference between our estimate of change detection capacity and those reported in the literature is color variability. In many of the previous and most cited studies (e.g. Alvarez & Cavanagh, 2004; Luck & Vogel, 1997), their authors used fixed sets of colors across all trials. In contrast, in our study, the set of used colors changed randomly from trial to trial. From other studies, it is known that observers can rather efficiently use stimulus regularities to inflate the capacity of VWM (Brady, Konkle, & Alvarez, 2009). Having a fixed color set, observers in the classical studies could also expand their useful memory set size showing higher capacities. However, this expansion could not be explained by pure VWM, it had something to do with some type of long-term memory as well. In our experiments, observers could not form any reliable long-term trace that would help them in any given trial, so they had to rely solely on a current working memory trace.

The idea that VWM capacity for objects is low because of the between-trial color variation gains important support from the continuous report data – both from our experiments and the existing literature. It can be seen that using exactly the same stimulation but a different report method we came to the same capacity estimate of about only two individual object colors. Importantly, other studies using the continuous report with subsequent mixture modeling show approximately the same estimates: while the probability of storing an item in memory is near 1.0 for one or two items (Zhang & Luck, 2008), it drops down to approximately .75-.83 for three items (Fougnie, Asplund, & Marois, 2010) that corresponds to the capacity around 2.25-2.49 items, which is fairly below the magic number 4 (Cowan, 2001). In these continuous report studies, the colors were also selected randomly in each trial. Therefore, it is possible that color regularity could be an important factor that led to higher capacity estimates in the previous change detection studies. If this is the case then the true VWM capacity for objects can be even more limited than was thought (Cowan, 2001; Luck & Vogel, 1997, 2013). Of course, further research is necessary to test the role of stimulus regularity in VWM capacity.

Sampling vs. exhaustive ensemble encoding

Another important question, which we addressed after reporting basically similar capacities and precision of object and ensemble VWM, was whether ensemble encoding can be provided by object encoding. If the observ-

er shows exactly the same performance in both conditions, is it possible that VWM always encodes a few objects within its limited capacity? Or does it enlarge the unit and encode the ensemble in its entirety, as if it encoded a single object in the object condition? We refer the first and the second hypotheses to as sampling and exhaustive encoding, respectively. The experimental distinction between these two hypotheses seems very important for our study, given the debate about sampling vs. exhaustive coding in ensemble perception (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Alvarez, 2011; Ariely, 2008; Chong et al., 2008; Maule & Franklin, 2016; Marchant et al., 2013; Myczek & Simons, 2008; Simons & Myczek, 2008; Utochkin & Tiurina, 2014).

In Experiment 4, we directly manipulated local samples that our observers were likely to encode with high priority, because we attracted their exogenous attention to those samples. We asked whether these samples would bias encoding towards ensembles whose representatives are in the samples and/or away from ensembles that are not represented in the sample. We found no evidence that sample representativeness has any effect on capacity or precision. Even when the observers did not pay exogenous attention to any item from a probed ensemble, they remembered this ensemble as efficiently as those whose representatives had been attended. This result allows us to rule out sampling as a potential explanation for equal VWM parameters in the object and ensemble conditions of Experiments 1-3. Even though attentional salience of the cued object sample (Jiang et al., 2002) could let the

members of that sample enter VWM with higher probability, this could only affect their encoding among other individual objects, when they are treated as individuals. However, what happens at the individual object level seems to keep ensemble coding intact. We conclude, therefore, that each color-defined ensemble is likely to be coded as a unitary chunk, and the efficiency of its encoding does not differ substantially from the efficiency of encoding the objects. It even occurs despite low “objecthood” of those ensembles: they had no such internal unity since the spatial organization of ensemble members was poor.

Conclusion

In our study we asked whether individual objects and ensembles (multiple objects with poor objecthood) are encoded as similar or different units in VWM. Our question was motivated by a discrepancy in the existing quantitative data about VWM capacity for objects and ensembles. Using the same

set of methods and a proper psychophysical control, we discovered that both objects and ensembles are in fact encoded with the same efficiency (perhaps, some of the previous capacity estimates for objects are inflated). Finally, using attentional manipulations we demonstrated that ensembles are encoded exhaustively and not as limited samples of individual objects. The finding that ensembles can be as strong VWM units as individual objects are in line with other recent claims (e.g. Huang, 2015). Moreover, our finding that attentional manipulations with individual objects did not affect VWM for ensembles suggest an idea that objects and ensembles can be two different representational levels of VWM (Brady & Alvarez, 2011; Brady et al., 2011).

Acknowledgements

The authors thank Maria Yurevich, Lilit Dulyan, Ekaterina Yudina, and Ekaterina Volchenkova for their assistance in data collection.

References

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Science*, *15*, 122–131.
- 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.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*, 7345–7350.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception and Psychophysics*, *70*, 1325–1326.
- Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception and Psychophysics*, *77*, 1116–1131.

- Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1440–1449.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.
- Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *Psychological Record*, *59*, 171–186.
- 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., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4–4.
- Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception and Psychophysics*, *70*, 1327–1334.
- Chong, S. C., & Treisman, A. M. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.
- Chong, S. C., & Treisman, A. M. (2005). Statistical processing: Computing average size in perceptual groups. *Vision Research*, *45*, 891–900.
- Corbett, J. E. (2017). The whole warps the sum of its parts: Gestalt-defined-group mean size biases memory for individual objects. *Psychological Science*, *28*(1), 12–22.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*, 3181–3192.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, R751–R753.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718–734.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576.
- Huang, L. (2015). Statistical properties demand as much attention as object features. *PLoS ONE*, *10*(8), e0131191.
- Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*, 663–676.
- Im, H. Y., Park, S. J., & Chong, S. C. (2015). Ensemble statistics as units of selection. *Journal of Cognitive Psychology*, *27*, 114–127.
- Jiang, Y., Chun, M. M., & Marks, L. E. (2002). Visual marking: selective attention to asynchronous temporal groups. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 717–730.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Science*, *17*, 391–400.
- Marchant, A. P., Simons, D. J., & De Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142*, 245–250.
- Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to an limited capacity subsampling mechanism. *Journal of the Optical Society of America A*, *33*, A22–A29.
- Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception and Psychophysics*, *70*, 772–788.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception and Psychophysics*, *64*, 1055–1067.
- Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception and Psychophysics*, *44*, 369–378.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, *16*, 283–290.
- Pierce, J. W. (2007). PsychoPy – psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.
- Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, *13*(8), 6. Retrieved from <http://www.journalofvision.org/content/13/8/6>
- Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(12), 18. Retrieved from <http://www.journalofvision.org/content/11/12/18>
- Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception and Psychophysics*, *70*, 1335–1336.
- Suchow, J. W., Brady, T. F., Fougner, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9.
- Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 331–351.
- Utochkin, I. S. (2016). Visual enumeration of spatially overlapping subsets. *The Russian Journal of Cognitive Science*, *3*, 4–20.
- Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, *146*, 7–18.
- Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual-system averages speed information. *Vision Research*, *32*, 931–941.
- Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). The efficiency of feature-based subitization and counting. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1449–1462.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast. *Psychonomic Bulletin and Review*, *18*, 484–489.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135.
- Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 601–621.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 452, 233–235.



Yuri A. Markov – research assistant, Laboratory for Cognitive Research, National Research University Higher School of Economics. Research area: cognitive psychology, visual attention, visual search, feature binding, visual working memory
E-mail: ymarkov@hse.ru



Natalia A. Tiurina – research fellow, Laboratory for Cognitive Research, National Research University Higher School of Economics, Ph.D. Research area: cognitive psychology, visual attention, visual search, memory, psychometrics
E-mail: ntyurina@hse.ru



Yulia M. Stakina – expert, Laboratory for Cognitive Research, National Research University Higher School of Economics, Ph.D. Research area: attention, individual differences
E-mail: stakina@hse.ru



Igor S. Utochkin – head, Laboratory for Cognitive Research; associate professor, Faculty of Social Sciences, School of Psychology, National Research University Higher School of Economics, Ph.D. Research area: visual attention, visual search, object and scene perception, ensemble summary statistics, visual working memory
E-mail: isutochkin@inbox.ru

Объем и точность зрительной рабочей памяти на объекты и ансамбли

Ю.А. Марков^а, Н.А. Тюрина^а, Ю.М. Стакина^а, И.С. Уточкин^а

^а *Национальный исследовательский университет «Высшая школа экономики», 101000, Россия, Москва, ул. Мясницкая, д. 20*

Резюме

Предыдущие исследования показывают, что объем зрительной рабочей памяти на цвета объектов ограничен и равен примерно 3–5 элементам. Данные, полученные в других исследованиях, утверждают, что множественные объекты могут храниться в форме более компактного представления — зрительного ансамбля. В то же время исследования зрительных ансамблей показывают, что в зрительной рабочей памяти одновременно могут храниться только два ансамбля. Возможно, данные различия связаны со способами измерения характеристик зрительной рабочей памяти, использованными в различных исследованиях: запоминание ансамблей не тестировалось при помощи методов, используемых для исследования памяти на объекты. Мы измерили объем и точность зрительной рабочей памяти для объектов и ансамблей, используя два стандартных метода — метод обнаружения изменений и метод градуального отчета с использованием моделей смешения (Mixture model). Мы обнаружили, что объем и точность зрительной рабочей памяти на объекты и ансамбли одинаковы, при контроле основных психофизических параметров: фовеальной плотности и площади предъявления ансамблей и объектов. Мы также показали, что сходство объемов зрительной рабочей памяти для объектов и ансамблей обеспечивается механизмом, позволяющим хранить ансамбль в зрительной рабочей памяти в форме целостной репрезентации, что аналогично эффективному хранению информации о единичных объектах.

Ключевые слова: зрительная рабочая память, восприятие объектов, восприятие ансамблей.

Марков Юрий Алексеевич — стажер-исследователь, научно-учебная лаборатория когнитивных исследований, Национальный исследовательский университет «Высшая школа экономики».

Сфера научных интересов: когнитивная психология, зрительное внимание, зрительный поиск, связывание признаков, зрительная рабочая память.

Контакты: umarikov@hse.ru

Тюрина Наталья Александровна — научный сотрудник, научно-учебная лаборатория когнитивных исследований, Национальный исследовательский университет «Высшая школа экономики», кандидат психологических наук.

Сфера научных интересов: когнитивная психология, зрительное внимание, зрительный поиск, психология памяти, психометрика.

Контакты: ntyurina@hse.ru

Стакина Юлия Михайловна — эксперт, научно-учебная лаборатория когнитивных исследований, Национальный исследовательский университет «Высшая школа экономики», кандидат психологических наук.

Сфера научных интересов: психология внимания, психология индивидуальных различий.
Контакты: stakina@hse.ru

Уточкин Игорь Сергеевич — заведующий лабораторией, научно-учебная лаборатория когнитивных исследований, Национальный исследовательский университет «Высшая школа экономики», кандидат психологических наук.

Сфера научных интересов: зрительное внимание, зрительный поиск, восприятие объектов и сцен, статистическая репрезентация ансамблей, зрительная рабочая память.
Контакты: isutochkin@inbox.ru