

Crowdsourced single-trial probes of visual working memory for irrelevant features

Hongsup Shin

Department of Neuroscience, Baylor College of
Medicine, Houston, TX, USA
Present address: Center for Neural Science and
Department of Psychology, New York University,
New York, NY, USA



Wei Ji Ma

Department of Neuroscience, Baylor College of
Medicine, Houston, TX, USA
Present address: Center for Neural Science and
Department of Psychology, New York University,
New York, NY, USA



We measured the precision with which an irrelevant feature of a relevant object is stored in visual short-term memory. In each experiment, 600 online subjects each completed 30 trials in which the same feature (orientation or color) was relevant, followed by a single surprise trial in which the other feature was relevant. Pooling data across all subjects, we find in a delayed-estimation task but not in a change localization task that the irrelevant feature is retrieved, but with much lower precision than when the same feature is relevant: The irrelevant/relevant precision ratio was 3.8% for orientation and 20.4% for color.

Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Wheeler & Treisman, 2002). If VSTM is object-based, then one could surmise that encoding a task-relevant feature of an object automatically causes irrelevant features of that object to be encoded as well (Hyun, Woodman, Vogel, Hollingworth, & Luck, 2009; Luria & Vogel, 2011; Shen, Tang, Wu, Shui, & Gao, 2013; Vogel et al., 2001; Yin et al., 2012).

This hypothesis has mostly been tested by examining whether the addition of an irrelevant feature decreases performance. Studies employing orientation-color change detection (Vogel et al., 2001) and color-shape change detection (Luria & Vogel, 2011) showed no effect, suggesting that people do not encode irrelevant features. However, these results can also be explained by the irrelevant feature having an independent pool of memory resource rather than sharing resources with the relevant feature. In fact, Hyun et al. (2009) found the opposite result: Subjects were more error-prone when one object changed in its irrelevant feature. When the authors introduced changes in all objects, they found an even stronger impairment, leading them to conclude that irrelevant features are encoded. Recent studies (Shen et al., 2013; Yin et al., 2012) found similar effects and interpreted them as evidence that VSTM is object-based.

Although the origin of the differences between the results of these studies remains unclear, even the positive results leave open the question of *how well* the irrelevant feature is stored in VSTM, and in particular if it is stored with the same precision as when that feature is relevant. To address these questions, it is

Introduction

Visual input originates from multiple objects that all have many features, such as orientation, color, and motion. To deal with this deluge of information, it is useful to have a short-term buffer—visual short-term memory (VSTM). VSTM for objects that have more than one feature has been an area of enduring interest in cognitive psychology. One prominent question has been whether multifeature objects get stored in VSTM as entire objects or as loose collections of features. This question has multiple aspects, one of which is whether all features of a task-relevant object are stored automatically, regardless of the relevance of each individual feature (Alvarez & Cavanagh, 2004; Bays, Wu, & Husain, 2011; Fougner, Asplund, & Marois, 2010; Jiang, Olson, & Chun, 2000; Lee & Chun, 2001;

Citation: Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5):10, 1–8, doi:10.1167/16.5.10.



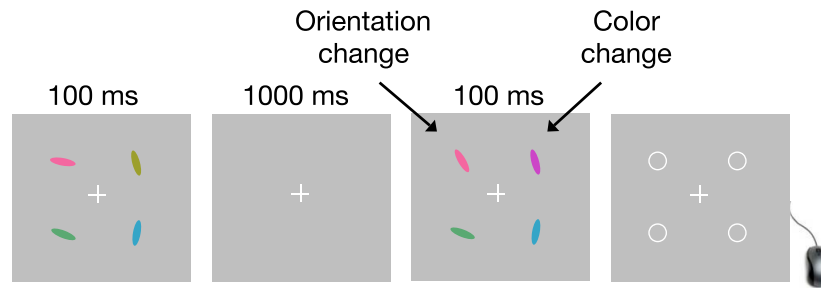


Figure 1. Trial procedure in the change localization experiments (laboratory experiments and Experiments 1 and 2). On each trial, there is both an orientation change and a color change. Subjects click on the location where a relevant-feature change occurred.

insufficient to only measure performance on trials in which the relevant feature is probed: Data must be collected on irrelevant-feature trials to make a comparison. This, however, brings about a problem: As soon as a subject experiences a surprise trial on which the irrelevant feature is probed, that feature becomes relevant. Therefore, each subject can be tested on only a single irrelevant-feature trial.

To solve this problem, we crowdsourced data from the Amazon Mechanical Turk, an online platform for data collection. We used stimuli that each had both an orientation and a color. We crossed two experimental paradigms (change localization and delayed estimation) with two options for which feature was irrelevant (orientation and color), for a total of four experiments. We found that people could recall the irrelevant feature, suggesting that it is encoded automatically.

Laboratory experiments

Methods

To set an appropriate level of difficulty, we conducted two laboratory experiments, each with five subjects. We used a change localization task (van den Berg, Shin, Chou, George, & Ma, 2012), noted in Figure 1; one experiment involved orientation change localization, the other color change localization. Subjects briefly viewed a sample display that consisted of four colored ellipses. The orientations and colors of the ellipses were drawn from uniform distributions, independently across items and features. After a delay, subjects viewed a test display, in which one ellipse (chosen with equal probabilities) changed orientation and another ellipse (independently chosen, also with equal probabilities) changed color. Both changes were independently drawn from uniform distributions. By chance, both changes could occur in the same object. One feature was relevant and the other was irrelevant. The task was to click on the location of the object that had changed in its relevant feature. At the beginning of

each session, we instructed subjects that their task was to localize the change in the relevant feature, that the change could be small or large, that one randomly chosen object would change in its irrelevant feature, and that by chance the same object could change in both features. Each experiment consisted of four blocks of 150 trials, for a total of 600 trials.

Stimuli were displayed on a 19" LCD monitor at a viewing distance of approximately 60 cm. Stimuli were presented on a midlevel grey background of luminance 33.1 cd/m^2 . Stimuli were equally spaced along an imaginary circle of radius approximately 7° of visual angle around fixation (calculated assuming a viewing distance of 60 cm), at angles $i/N \cdot 360^\circ$, where $i = 1, \dots, N$, and $N = 4$. The experiments were programmed using Psychophysics Toolbox in MATLAB (Brainard, 1997; Pelli, 1997).

Results

Average subject performance was above chance in both the orientation (black curve in Figure 2A; right-tailed z test for proportions on data pooled across subjects: $z = 73.9$, $p < 0.001$) and color experiments (black curve in Figure 2B; $z = 83.5$, $p < 0.001$). A logistic regression revealed a significant main effect of change magnitude in the orientation experiment ($r = 0.023 \pm 0.001$, $p < 0.001$) and in the color experiment ($r = 0.026 \pm 0.001$, $p < 0.001$). These results indicated that this paradigm might be suitable for the crowdsourced experiments.

Experiment 1

Methods

The experiment was conducted online on www.mturk.com. Subjects enrolled by selecting our experiment from a list of "Human Intelligence Tasks." To maximize consistency in response modality and viewing

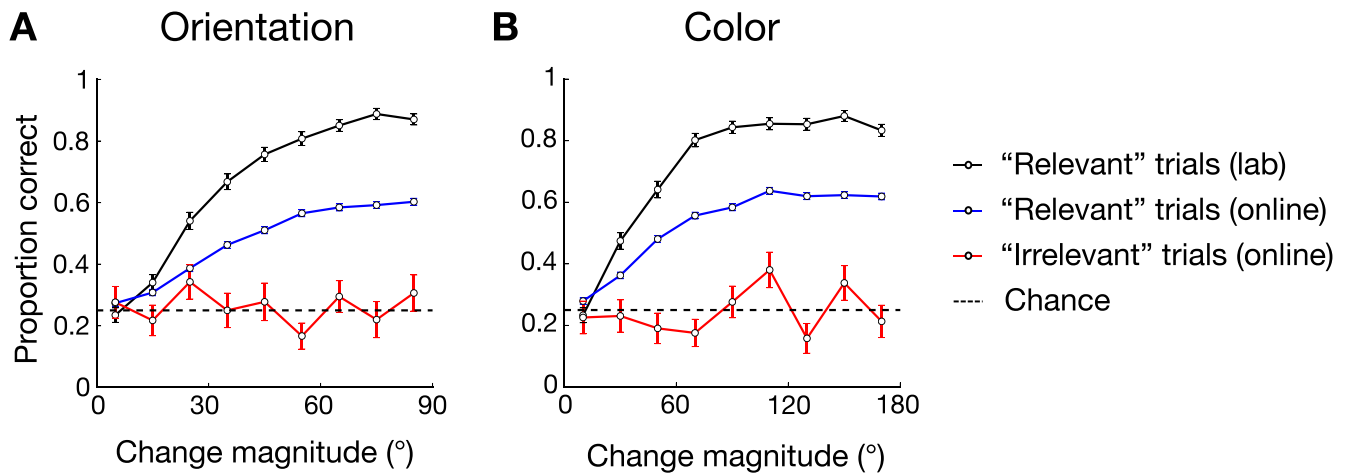


Figure 2. Performance in the change localization experiments. Proportion correct as a function of change magnitude (binned for plotting purposes). Per bin, we pooled data across subjects. (A) Orientation: black—laboratory experiment (five subjects, 600 trials per subject), blue—irrelevant-feature trials of online subjects (600 subjects, 30 trials per subject), red—irrelevant trials of online subjects (600 subjects, one trial per subject). Error bars are 68% confidence intervals for binomial proportions. Performance on the irrelevant trials was indistinguishable from chance. (B) As (A), but for color change localization. Again, performance on the irrelevant-feature trials was indistinguishable from chance.

conditions, we prevented enrollment via mobile devices such as smartphones or tablets. Once a subject selected the experiment, they were told to enlarge the window size of their web browser to at least 800×600 pixels. If they did not do this, they could not continue. Then, the subject completed a six-item Ishihara Color Test to test for color vision deficiencies. A subject who passed this test was then led step-by-step through an example trial accompanied by on-screen instructions. This instruction phase was self-paced and subjects could freely move back and forth between the screens of the example trial. Taking this selection into account, 600 subjects participated. Each experiment lasted approximately 5 min, and each subject was paid 25 cents. The experimental protocol was approved by the Institutional Review Board of Baylor College of Medicine.

Similar to those from the orientation laboratory experiment, the stimuli were four colored ellipses (Figure 1), presented in a rectangular 400×400 -pixel, midlevel grey window. The ellipse had minor and major axes were 16 and 7 pixels, respectively. The colors were drawn from 360 values uniformly distributed along a circle in the fixed- L^* plane of CIE 1976 (L^* , a^* , b^*) color space corresponding to a center $(a^*, b^*) = (12, 13)$ and radius 60. We could not control luminance. Stimuli were equally spaced along an imaginary circle around fixation, at angles $i/N \cdot 360^\circ$, where $i = 1, \dots, N$, and $N = 4$. The experiments were programmed in JavaScript, HTML and CSS.

“Relevant” trials

The experiment consisted of thirty “relevant” trials followed by one “irrelevant” trial (Figure 3A, top).

The sequence of a relevant trial consisted of the presentation of a fixation cross (1000 ms), the first stimulus array (150 ms), a delay period in which the fixation cross was present (1000 ms), the second stimulus array (150 ms), and a response screen (present until response). The second array was identical to the first except that one randomly chosen object was different in its relevant feature (orientation) value from the first array, and another randomly chosen object in its irrelevant feature (color). The response screen included a text message in the center, reminding a subject of the task (“Where was the ORIENTATION change?”). The magnitude of the change was independently drawn from a uniform distribution for each feature. For each feature, each object had an equal probability of changing. Thus, by chance, changes in the irrelevant and relevant feature could occur in the same object. The response screen consisted of empty circles at the same locations as where objects were presented in the stimulus arrays. The task was to click on the location of the object that had changed its orientation. After the response, feedback was provided: The fixation cross turned green if the response was correct, red if the response was incorrect. Before subjects started the experiment, they completed a demonstration trial that reflected the trial procedure of the relevant trials. This trial consisted of static images of the displays, accompanied by written descriptions.

“Irrelevant” trial

The last trial (31st) was identical to an irrelevant trial except for the following: After subjects clicked the

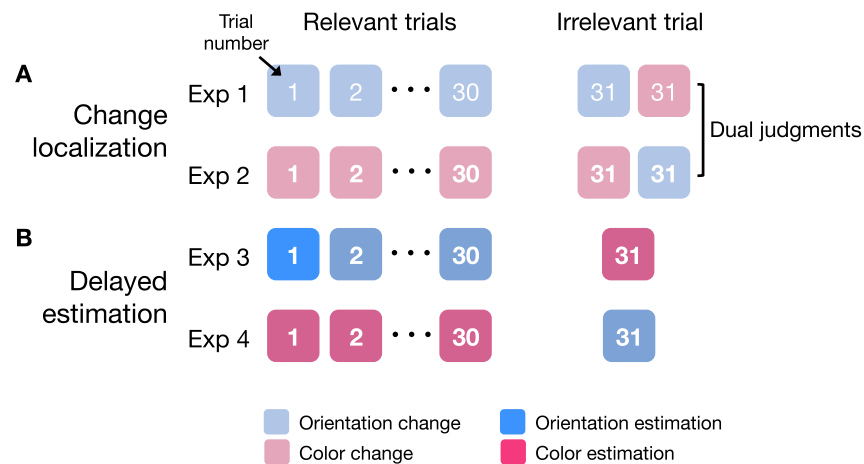


Figure 3. Online experiments. (A) Experiments 1 and 2 featured change localization tasks. Experiment 1: An irrelevant color trial followed 30 orientation-relevant trials. On the irrelevant trial, subjects were first asked to report the location of the orientation change, and then to report the location of the color change. Experiment 2: Like Experiment 1, except that color was relevant and orientation was irrelevant. (B) Experiments 3 and 4 featured delayed-estimation tasks. Experiment 3: An irrelevant color trial followed 30 orientation-relevant trials. On the irrelevant trial, subjects were asked to report the color of the stimulus on a color wheel. Experiment 4: Like Experiment 3, except that color was relevant and orientation was irrelevant.

location of relevant-feature change at the response screen, another response screen appeared with four empty circles and a message saying, "... and where was the COLOR change?" After subjects clicked one of the circles, the experiment ended.

Results

We pooled the data across all subjects. On the *relevant trials*, subjects' average performance was significantly above chance (blue curve in Figure 2A, z test for proportions, $z = 69.1$, $p < 0.001$). To compare performance between Mechanical Turk and lab subjects, we conducted a logistic regression with subject group (0: lab; 1: Mechanical Turk) and change magnitude as regressors. The coefficient for subject group was estimated at -0.78 ± 0.03 and was highly significant ($p < 0.001$), indicating that Mechanical Turk subjects performed worse than lab subjects. On the *irrelevant trials*, performance was not significantly different from chance (red curve in Figure 2B, 0.243 ± 0.018 ; $z = 0.66$, $p = 0.26$).

Experiment 2

Methods

Experiment 2 was identical to Experiment 1 except that we switched orientation and color. Thus, the goal was to examine whether a subject could retrieve an

irrelevant orientation memory. To recruit naïve subjects, we blocked subjects who participated in Experiment 1 by blocking their user IDs. Thus, 600 different subjects participated in Experiment 2.

Results

The results were very similar to Experiment 1. We again pooled the data across all subjects. On the *relevant trials*, subjects' average performance was significantly above chance (blue curve in Figure 2B, z test for proportions, $z = 86.5$, $p < 0.001$). The logistic regression coefficient for subject group was estimated at -1.00 ± 0.03 and was highly significant ($p < 0.001$), indicating that Mechanical Turk subjects performed worse than lab subjects. On the *irrelevant trials*, performance was not significantly different from chance (red curve in Figure 2B, 0.261 ± 0.018 ; $z = -0.19$, $p = 0.43$).

Discussion

Several factors might have contributed to the chance performance on irrelevant-feature trials in Experiments 1 and 2. First, we asked subjects first to locate the relevant change, and only afterwards to locate the irrelevant change. Thus, the subject's report of the irrelevant feature had a longer delay than when that feature was relevant, potentially decreasing performance. Second, subjects might discard their memory of the irrelevant feature as soon as they believe the trial to

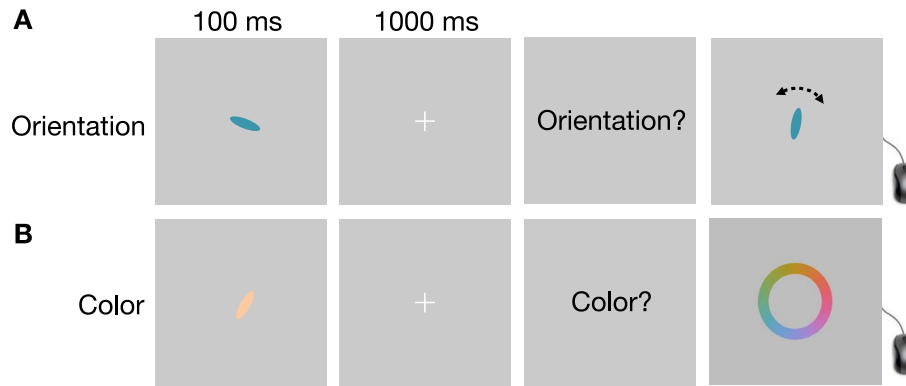


Figure 4. Trial procedure in Experiments 3 and 4 (delayed estimation). (A) Orientation estimation trial. Subjects estimated the orientation of the stimulus by rotating the probe using the mouse. (B) Color estimation trial. Subjects estimated the color of the stimulus by clicking on the color wheel.

be finished, which is right after their relevant-feature report. This would be consistent with a study that found that the encoding of features in multifeature objects is obligatory, but that maintenance is voluntary (Marshall & Bays, 2013). Second, subjects might not have fully understood that they had to do something different for the second report on the last trial than for the first report. Indeed, on the irrelevant-feature trial, 36% (Experiment 1) and 41% (Experiment 2) of subjects clicked twice on the same location; both proportions are significantly higher than chance (z test for proportions; $p < 10^{-9}$). It is possible that subjects were confused, because the response modality was identical for both judgments, namely clicking on one of four empty circles. Fourth, it is possible an irrelevant feature is encoded at smaller set sizes but not at larger ones.

Experiment 3

We identified several problems that could have produced the null results of Experiments 1 and 2: a longer delay, discarding information after a decision, a misunderstanding of the instructions, and set size. To address these problems, we conducted two more experiments, Experiments 3 and 4, in which we used a different experimental paradigm, called delayed estimation (Wilken & Ma, 2004). Subjects viewed a brief display consisting of a single stimulus, and, after a delay, were asked to estimate the feature value of a randomly selected stimulus on a continuum. This removed or reduced the problems above: (a) the subject does not make a decision before they are asked about the irrelevant feature; (b) the delay between the offset of the memory array and the subject's response is similar on the irrelevant and the relevant trials; (c) the response modality was entirely different on the

irrelevant trial than on the relevant trials—for example, subjects clicked on a color wheel instead of rotating an orientation probe; (d) set size was 1.

Methods

Experiment 3 was identical to Experiment 1 except for the following. Set size was 1, and the ellipse appeared in the center of the window. On the first 30 (relevant-feature) trials (Figure 4A), the response stage consisted of a response probe appearing with a message asking subjects to report the orientation of the memorized ellipse. The probe was a colored ellipse with the same color as the stimulus and an orientation drawn from a uniform distribution. Subjects could rotate the ellipse by moving a mouse, and submitted their response by pressing the spacebar. Feedback was then provided by presenting the original stimulus (“correct”) and the reported stimulus (“reported”) simultaneously in the top and bottom parts of the stimulus window, respectively. On the 31st (irrelevant-feature) trial (Figure 4B), after the delay period, a color wheel appeared with a message asking subjects to report the color of the stimulus by clicking on the color wheel. No feedback was given on this trial, and the experiment ended immediately afterwards. Again, 600 new subjects were recruited for this study.

Results

The pooled distribution of the estimation error on the relevant-orientation trials was significantly different from uniform (Kolmogorov-Smirnov test: $D = 0.35$, $p < 0.001$, Figure 5A). On a $[-90^\circ, 90^\circ]$ domain, the error distribution on relevant-orientation trials had a circular mean of -0.3° and a circular standard deviation

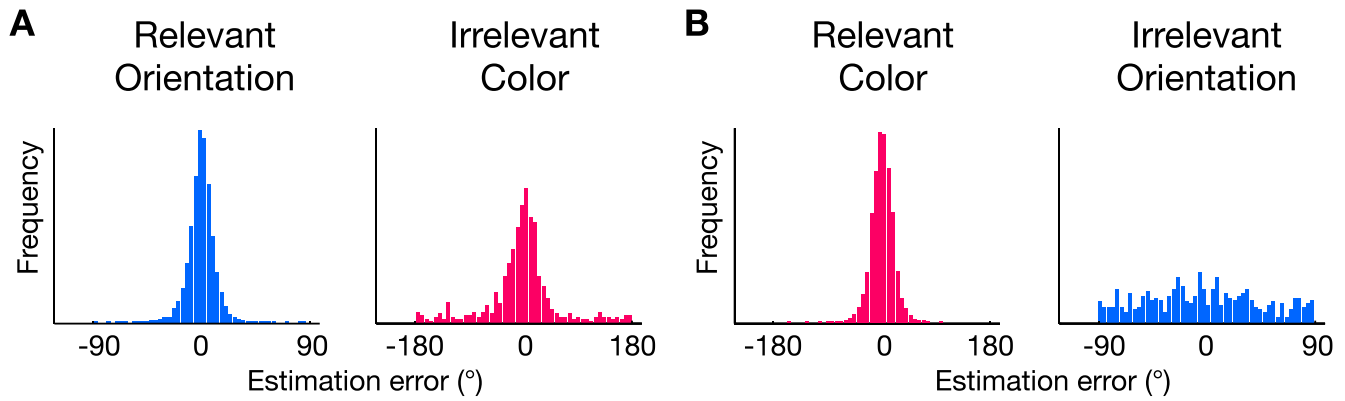


Figure 5. Performance in the delayed-estimation experiments. In each experiment, 600 subjects each performed 30 relevant-feature trials and one irrelevant-feature trial. Thus, the relevant-feature histograms are based on 18,000 trials and the irrelevant-feature histograms on 600 trials. (A) Experiment 3. Orientation was relevant and color irrelevant. Both distributions are significantly different from uniform. (B) Experiment 4. Color was relevant and orientation was irrelevant. Both distributions are significantly different from uniform.

(Mardia & Jupp, 1999) of 13.7° , the latter with a 95% bootstrapped confidence interval of $[13.4^\circ, 14.0^\circ]$. Critically, the error distribution on irrelevant-color trials was not uniform either ($D = 0.28$, $p < 0.001$, Figure 5A). On a $[-180^\circ, 180^\circ]$ domain, it had a circular mean of -3.3° and a circular standard deviation of 51.6° , the latter with a 95% bootstrapped confidence interval of $[47.4^\circ, 56.0^\circ]$. This suggests that subjects were able to retrieve the irrelevant color.

Experiment 4

Methods

Experiment 4 was identical to the first experiment except that the first 30 trials were relevant-color trials (Figure 4B) and the surprise trial was a relevant-orientation trial (Figure 4A). Again, 600 new subjects were recruited for this study.

Results

Results were similar to those of Experiment 3. The pooled distribution of the estimation error on the relevant-color trials was significantly different from uniform ($D = 0.36$, $p < 0.001$, Figure 5B). On a $[-180^\circ, 180^\circ]$ domain, the error distribution on relevant-color trials had a circular mean of -0.5° and a circular standard deviation of 23.3° , the latter with a 95% bootstrapped confidence interval of $[22.8^\circ, 23.8^\circ]$. Critically, the error distribution on irrelevant-orientation trials was not uniform either ($D = 0.09$, $p < 0.001$, Figure 5B). On a $[-90^\circ, 90^\circ]$ domain, it had a circular

mean of -2.3° and a circular standard deviation of 70.6° , the latter with a 95% bootstrapped confidence interval of $[60.6^\circ, 83.0^\circ]$. This shows that subjects were able to retrieve irrelevant working memories of orientation.

We now examine the circular standard deviations from Experiments 3 and 4 together. We define an irrelevant/relevant precision ratio (IRPR) as:

$$\text{IRPR} = \left(\frac{\text{circular } SD \text{ when feature is irrelevant}}{\text{circular } SD \text{ when feature is relevant}} \right)^2.$$

We find an IRPR of 3.8% for orientation and 20.4% for color. The fact that both are below 100% indicates that memory precision is lower when a feature is irrelevant than when the same feature is relevant. The difference between these ratios could have several possible causes. First, the response modality differed between orientation and color: On an orientation trial, subjects rotated the orientation probe, which only allowed them to see one orientation at a time. However, in a color trial, subjects could see the entire color wheel, providing a continuum of templates for comparison to the memory. Second, rotating the orientation probe took slightly more time than clicking on the color wheel, possibly reducing the quality of the memory. Third, color might ecologically be more important than orientation, and therefore be encoded better when not explicitly relevant.

General discussion

We used crowdsourcing to measure the precision of VSTM encoding for irrelevant features. In a change

localization task, performance on irrelevant-feature trials was at chance, but this might have been due to confounds in the experimental design. Using a simpler and more direct design, we found that subjects were capable of retrieving the irrelevant feature, although with substantially lower precision than the relevant feature.

A recent study (Chen & Wyble, 2015) asked a question similar to ours, but did not use crowd-sourcing. Subjects viewed a brief display consisting of three numbers and one letter, each with a different color. For the first 155 (or 11) trials, subjects located the letter. Then, on the 156th (or the 12th) trial, the experimenters surprised the subjects by asking them to recall the identity or the color of the letter, rather than its location. The authors found that the proportion of the subjects who correctly answered the surprise trial was close to chance level (25%), and concluded that people failed to recall irrelevant features. A problem with this study is that they based this conclusion on 20 subjects and therefore a total of 20 useful trials; This provides very limited statistical power.

Earlier studies examined irrelevant-feature memory indirectly by comparing change detection performance in the presence and absence of an irrelevant feature (Luria & Vogel, 2011; Vogel et al., 2001). The authors found that the presence of an irrelevant feature did not affect performance. However, as in the change localization task in our Experiments 1 and 2, change detection consists of an encoding and a decision-making stage (Keshvari, van den Berg, & Ma, 2012). In the encoding stage, stimuli are represented as noisy memories. Then, observers make a decision on the occurrence of a change based on the measurements. Thus, Luria and Vogel's results could be explained by the irrelevant feature not being encoded, or by it being encoded but appropriately ignored in the decision stage. Our results favor the latter interpretation.

Our main results are limited to a set size of 1, a delay period of 1 s, 30 relevant-feature trials preceding the irrelevant-feature trial, and to combinations of orientation and color. It remains to be seen how the irrelevant/relevant precision ratio (IRPR) will change as these choices are varied. There is evidence that set size and delay period matter: In a color memory task, an fMRI study found that irrelevant shape affected activity in a brain area involved in shape representation during the delay period, but that this activity was short-lived and lower at higher set size (Xu, 2010). Our paradigm allows for a systematic characterization of the dependence of IRPR on set size, delay period, number of relevant-feature trials, and feature combination.

Finally, it might be interesting to determine whether the change localization paradigm we used in Experiments 1 and 2 can be modified such that the precision of the irrelevant-feature memory becomes measurable. For example, one could use a change detection task at set size 1, in which only the irrelevant feature is probed on the surprise trial.

Keywords: visual working memory, visual short-term memory, delayed estimation, object memory

Acknowledgments

This research was supported by NIH grant R01EY020958 to W. J. M.

Commercial relationships: none.

Corresponding author: Wei Ji Ma.

Email: weijima@nyu.edu.

Address: Center for Neural Science and Department of Psychology, New York University, New York, NY, USA.

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.
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Chen, H., & Wyble, B. (2015). Amnesia for object attributes: Failure to report attended information that had just reached conscious awareness. *Psychological Science*, *26*(2), 203–210.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12):27, 1–11, doi:10.1167/10.12.27. [PubMed] [Article]
- Hyun, J., Woodman, G. F., Vogel, E. K., Hollingworth, A., & Luck, S. J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1140–1160.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal*

of *Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 683–702.

- Keshvari, S., van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, 7(6), e40216.
- Lee, D., & Chun, M. 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. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, 49(6), 1632–1639.
- Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. Chichester, UK: John Wiley & Sons.
- Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, 13(2):21, 1–13, doi:10.1167/13.2.21. [PubMed] [Article]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Shen, M., Tang, N., Wu, F., Shui, R., & Gao, Z. (2013). Robust object-based encoding in visual working memory. *Journal of Vision*, 13(2):1, 1–11, doi:10.1167/13.2.1. [PubMed] [Article]
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA*, 109(22), 8780–8785.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48–64.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12):11, 1120–1135, doi:10.1167/4.12.11. [PubMed] [Article]
- Xu, Y. (2010). The neural fate of task-irrelevant features in object-based processing. *Journal of Neuroscience*, 30(42), 14020–14028.
- Yin, J., Zhou, J., Xu, H., Liang, J., Gao, Z., & Shen, M. (2012). Does high memory load kick task-irrelevant information out of visual working memory? *Psychonomic Bulletin & Review*, 19(2), 218–224.