Supporting online material

Two items remembered as precisely as one: How integral features can improve visual working memory

Gi Yeul Bae & Jonathan I. Flombaum
Johns Hopkins University

Contents:

Additional Analyses S1 (Model with guessing)
Table S2 (Parameter values)
Additional Analyses S3 (Edge effects)
Figure S4 (Schematic depiction of consistency analysis)
Figure S5 (Consistent vs. Inconsistent Error rates)

Additional Analyses S1: mixture models

In the main text, we report estimated precision derived from a probit regression model fit to the results of each experiment and condition. Many recent studies, however, utilize a mixture modeling approach, under the assumption that participants guess randomly, at times, and that such guessing may reflect capacity limits in memory (e.g. Zhang & Luck, 2008). To ensure that our precision estimates were not unduly contaminated by guessing, we additionally fit a mixture model to the results of Experiments 3-5.

It should be noted that the current study utilized binary decisions. So the two components of a mixture model employed here are a psychometric function, typically assumed to be a normal CDF, and a uniform guessing rate.

Thus we modeled the results assuming that the probability of a response (*r*) (brighter, bigger, or louder), given the magnitude of the change between the target and the probe, follows a mixture of a normal CDF and a uniform guessing distribution. Below, the parameter P_{mem} designates the likelihood that a response was drawn from a normally distributed memory representation, and in turn 1-P_{mem} designates the likelihood of random guess responses.

$$P(r \mid x) = P_{mem} * \mathit{normcdf}\,(x, SD) + (1\text{-}P_{mem}) * \frac{1}{2} \quad (1)$$

Here *x* is the magnitude of the difference between the probe and the target item, and SD is the standard deviation of the normal CDF. The best set of two free parameters, SD and P_{mem} , were obtained by minimizing the mean of the squared differences between empirical observations and model predictions with maximum allowed iterations of 10,000.

$$MSE = \frac{1}{n}\Sigma(\text{observation} - \text{prediction})^2 \qquad (2)$$

Table S1 reports the SD and P_{mem} values obtained.

		One Item	SIF	DIF
Experiment 3 (Luminance)	SD	0.065	0.098	0.065
	P_{mem} (%)	93.71	94.45	93.71
Experiment 4 (Size)	SD	0.104	0.205	0.101
	P_{mem} (%)	92.40	91.72	91.44
Experiment 5 (Amplitude)	SD	0.161	0.195	0.129
	P_{mem} (%)	99.99	99.99	92.51

Table S2. Results of mixture modeling. Two free parameters (SD and P_{mem}) were obtained for each experiment and condition. Overall, P_{mem} values were high, and did not differ greatly by condition. SD values followed the same patterns as reported in the main text on the basis of a probit regression model (note, higher SD values imply less precision).

Additional Analyses S3: Edge effects

Studies of visual working memory frequently employ a delayed estimation task with a circular response space (e.g. Zhang & Luck, 2008). However, we used a discrimination task and a non-circular set of features. Such tasks are known to be influenced by 'edge effects' (Thiele et al., 2011). That is, if a probe value is near the edge of the range of potential values, directional responses in the direction of that edge should be more probable. Thus unexpectedly good performance may be driven by high confidence and high accuracy responses to probes near edges. (In the reported experiments, these would be very dim or very bright probes, very large or very small ones, or very loud or very soft ones). For the purposes of this study, this would be a problem if edge effects manifested themselves in a more pronounced way in trials with different integral features (DIF) —where performance was good— compared to trials with the same integral features (SIF) —where performance was worse.

Accordingly, we investigated whether edge effects were present, in the first place, and whether they differed across experimental conditions. Towards this end, we sorted trials based on the feature of the probe. We identified the 40% of probe values with the most extreme features on either end (20% on each end). Then, for each experiment (3-5) we conducted a two-way ANOVA with experimental condition (1 object, SIF, DIF) and edge condition (extreme 40% or middle 60%) as within subject factors.

For Experiment 3, luminance, the main effect of edge was significant in all three conditions, $F(1,11)= 268$, $p <.001$. (Accuracy for edge trials was higher than middle trials, $p <.001$ for all experimental conditions). Crucially, the interaction between experimental condition and edge condition was not significant, $F(2,22) = 2.65$, $p = 0.09$. And a planned comparison between the magnitude of the edge effect (accuracy in extreme minus middle trials) in SIF and DIF trials was not statistically significant ($t(11) = 0.85$, $p = 0.41$).

For Experiment 4, size, neither the overall edge effect, $F(1,7) = 1.55$, $p = 0.253$, nor the two-way interaction, $F(2,14) = 1.104$, $p = 0.359$, was significant. And a planned comparison showed that the magnitude of the edge effect was not different between SIF and DIF trials ($t(7) = .58$, $p = 0.55$).

For Experiment 5, the Volume experiment, the overall edge effect was significant, $F(1,11) = 78.3$, $p <.001$. But again, the interaction between edge condition and experimental condition was not statistically significant, $F(2,22) = 0.318$, $p =0.731$. A planned comparison showed that the magnitude of the edge effect in SIF trials was not different from DIF trials, $t(11) = 1.10$, $p = .2945$.

Overall, therefore, an edge effect was present in two from among the three relevant experiments. But in none of the experiments was there evidence of an interaction between edge effects and the critical experimental manipulations.

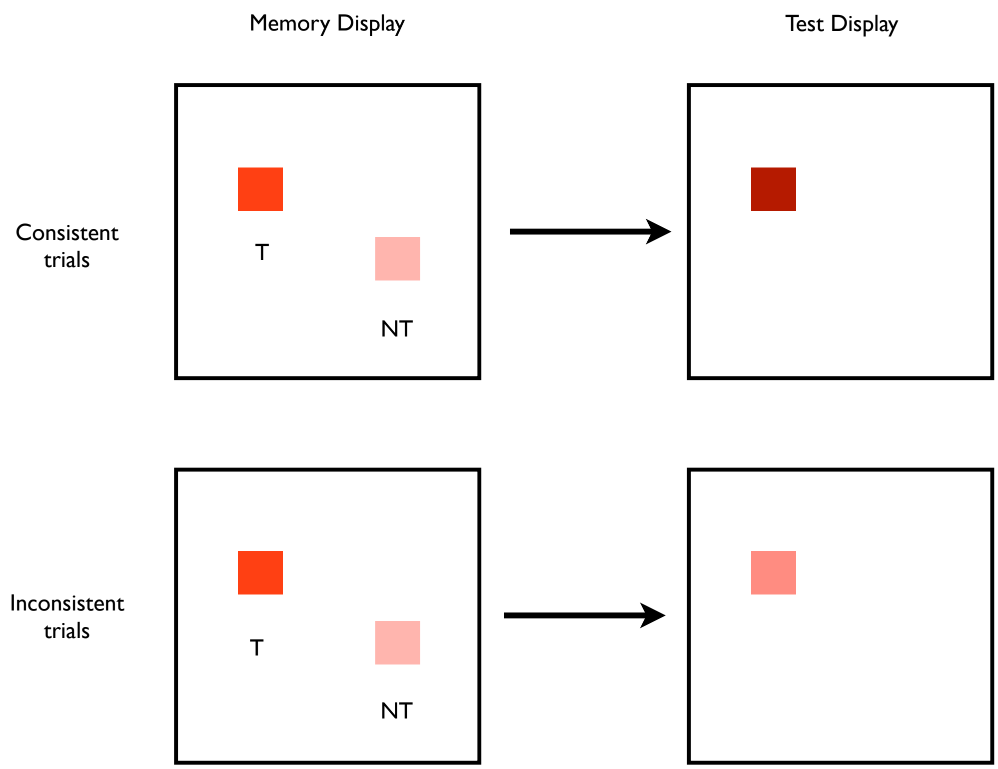

Figure S4. **Schematic depiction of Consistent/Inconsistent trial classification for Same Integral Feature trials in Experiment 3 (luminance working memory).** We classified trials on the basis of the response demanded by the nontarget item relative to the response demanded by the target item. In Consistent trials, both the target and nontarget demanded the same directional response —e.g. 'darker' in the schematic version shown. In Inconsistent trials, the nontarget driven response would be different from a target driven one —e.g. 'darker' in the schematic version shown, whereas the response to the target would be 'brighter.' The same logic was applied to classifying trials in Experiments 4 and 5 (Size and Amplitude).

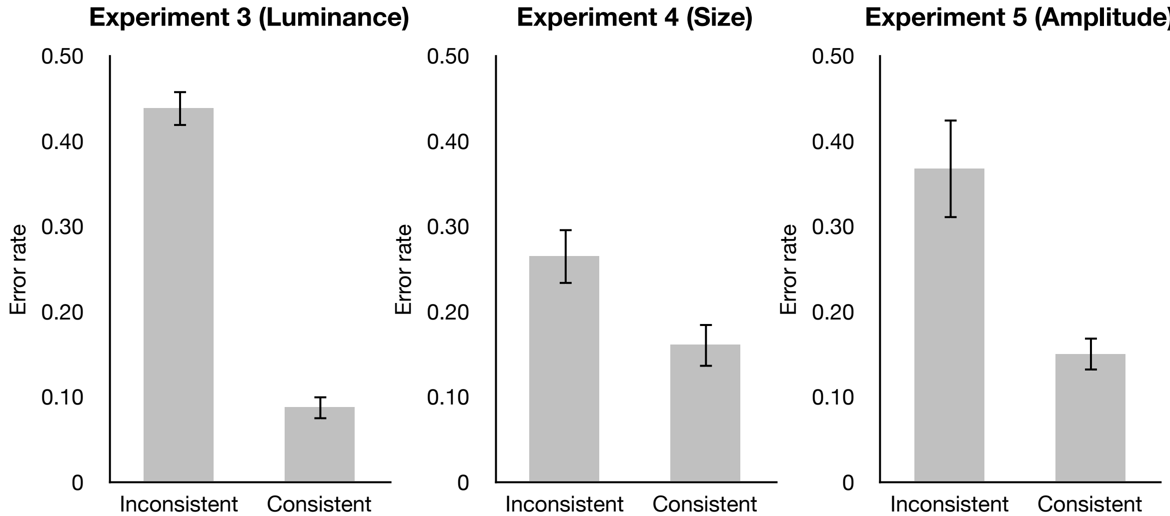

Figure S5. **Error rates in Consistent and Inconsistent trials (Same Integral Feature condition, Experiments 3-5)**. Error rates were higher in Inconsistent compared with Consistent trials (statistics reported in main text; error bars reflect ±1 S.E of the mean).

References

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

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