No effect of monetary reward in a visual working memory task

Ronald van den Berg[1,2], Qijia Zou[3], Wei Ji Ma[3,4]

[1]Department of Psychology, University of Uppsala, von Kraemers allé 1E, Uppsala, Sweden
[2]Department of Psychology, University of Stockholm, Frescati Hagväg 9A, Stockholm, Sweden
[3]Department of Psychology, New York University, 6 Washington Place, New York, NY, USA
[4]Center for Neural Science, New York University, 4 Washington Place, New York, NY, USA

ABSTRACT

Previous work has shown that humans distribute their visual working memory (VWM) resources flexibly across items: the higher the importance of an item, the better it is remembered. A related, but much less studied question is whether people also have control over the *total* amount of VWM resource allocated to a task. Here, we approach this question by testing whether increasing monetary incentives results in better overall VWM performance. In two experiments, subjects performed a delayed-estimation task on the Amazon Turk platform. In both experiments, four groups of subjects received a bonus payment based on their performance, with the maximum bonus ranging from $0 to $10 between groups. We found no effect of the amount of bonus on intrinsic motivation or on VWM performance in either experiment. These results suggest that resource allocation in visual working memory is insensitive to monetary reward, which has implications for resource-rational theories of VWM.

INTRODUCTION

A central question in research on human visual working memory (VWM) is how much flexibility exists in how the system distributes its resource across encoded items (Luck & Vogel, 2013; Ma, Husain, & Bays, 2014). The answer to this question partly depends on how one conceptualizes the nature of VWM resource. One class of models postulates that VWM consists of a small number of "slots" that each provide an indivisible amount of encoding resource (e.g., (Awh, Barton, & Vogel, 2007; Cowan, 2001; Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008)). Since the number of slots is typically assumed to be very small (3 to 4), these models allow for virtually no flexibility in resource allocation. A competing class of models conceptualizes VWM as a continuous resource (e.g., (Bays & Husain, 2008; Fougnie, Suchow,

& Alvarez, 2012; Keshvari, van den Berg, & Ma, 2013; Shaw, 1980; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004)), sometimes in combination with a limit on the number of encoded items (Sims, Jacobs, & Knill, 2012; van den Berg, Awh, & Ma, 2014). Since a continuous resource can be divided into arbitrarily small packages, these models allow for a high degree of flexibility in resource allocation.

Several recent studies have found evidence for flexibility in VWM resource allocation. First, it has been found in multiple experiments that when one item in a stimulus array is more likely to be selected for test than other items, subjects remember this item with better precision (Bays, 2014; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Emrich, Lockhart, & Al-Aidroos, 2017; Gorgoraptis, Catalao, Bays, & Husain, 2011; Yoo, Klyszejko, Curtis, & Ma, 2018; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011). In addition, it has been reported that subjects can make a tradeoff between the number of items in VWM and the quality with which they are encoded (Fougnie, Cormiea, Kanabar, & Alvarez, 2016; however see Zhang & Luck, 2011). The kind of flexibility found in these studies typically improves task performance compared to what can be achieved using a fixed allocation strategy, which suggests that the allocation is driven by a rational policy.

We recently formalized this suggestion by modeling VWM as a rational system that balances the amount of invested resource against expected task performance: the more there is at stake, the more resource is allocated for encoding (van den Berg & Ma, 2018). This "resource-rational" interpretation of VWM predicts two kinds of flexibility in the allocation of VWM resource. First, items of unequal importance are assigned unequal amounts of encoding resource, which is consistent with the findings cited above. Second, *tasks* of unequal importance are assigned unequal amounts of *total* resource: the higher the incentive to perform well on a task, the more VWM resource a subject should be willing to invest. In support of the second kind of flexibility, it has been found that subjects who are encouraged to "try to remember all items" in a change detection task have higher estimated numbers of slots than subjects who are told to "just do your best" or to "focus on a subset" (Bengson & Luck, 2016). Moreover, in one of our own studies, we observed that the estimated total amount of invested VWM resource in delayed-estimation tasks often varies non-monotonically with set size, in a way that can be explained by a resource-rational model (van den Berg & Ma, 2018). Finally, it has been reported that cueing can increase net VWM capacity (Myers, Chekroud, Stokes, & Nobre, 2018).

In the present study, we examine whether the total amount of allocated VWM resource is affected by monetary reward. We performed two experiments in which subjects earned a performance-contingent monetary bonus on top of a base payment. When encoding is costly, a rational observer should adjust its total amount of invested VWM resource to the amount of performance-contingent bonus: the higher the potential bonus, the more effort should be put into the task. In both experiments, we found no evidence for such an effect. In opposition to the prediction following from a resource-rational theory of VWM (van den Berg & Ma, 2018), the present results suggest that VWM resource allocation is insensitive to monetary reward.

EXPERIMENT 1

Data and code availability

All data, Matlab analysis scripts to reproduce figures of results, and JASP files with statistical analyses are available at https://osf.io/mwz27/.

Recruitment

Subjects were recruited on the Amazon Mechanical Turk platform, where the experiment was posted as a "Human Intelligence Task". The experiment was visible only to subjects who were located in the USA, had not participated in the experiment before, and had an approval rate of 95% or higher. A total of 355 subjects signed up, of which 156 were disqualified due to failing the post-instruction quiz (see below). The remaining 199 subjects were randomly assigned to four groups (n=49, 47, 47, 46) that differed in the total amount of bonus they could earn by performing well ($0, $2, $6, $10). Besides the bonus, subjects received a $1 base payment. The experiment was approved by the Institutional Review Board of New York University.

Stimuli and task

On each trial, the subject was presented with 1, 2, 4, 6, or 8 Gabor patches, which were placed along an invisible circle around a central fixation point (Fig. 1A). We refer to the number of presented items as the set size, which varied from trial to trial in a pseudo-random manner. The orientation of each patch was drawn independently from a uniform distribution over all possible orientations. The stimulus appeared for 50 milliseconds and was followed by an empty screen with a duration of 1 second (memory period). Thereafter, a randomly oriented Gabor patch appeared at one of the previous stimulus locations, whose initial orientation was randomly drawn and could be adjusted through mouse movement. The task was to match the orientation of this probe stimulus with the remembered orientation at that location. After submitting the response, the error between the correct orientation and the reported orientation, ε, was converted into an integer score between 0 and 10, with more points assigned for smaller errors (see

Appendix for a visualization of the scoring function). Feedback was provided after each trial by showing the obtained score and two lines that corresponded to the correct and responded orientations.

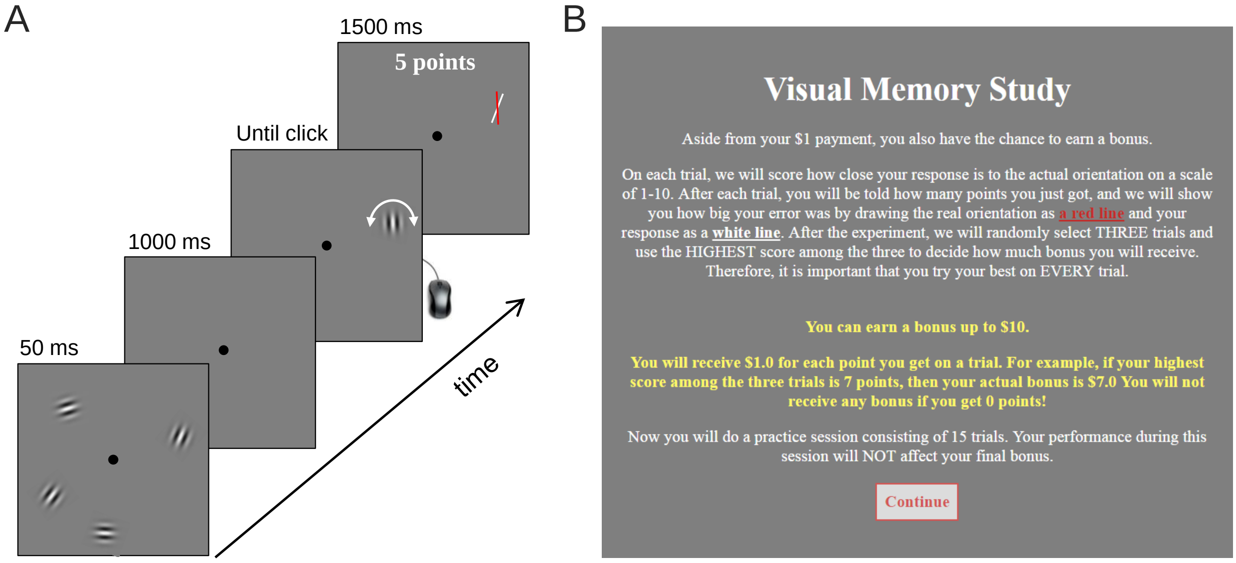

Figure 1 | Experimental procedure. (A) Illustration of a single trial in Experiment 1 (not to scale). Subjects were briefly presented with 1, 2, 4, 6, or 8 Gabor patches, which they had to keep in memory during the delay period. Thereafter, a randomly oriented Gabor patch would appear at one of the previous stimulus locations. The task was to match the orientation of this stimulus with the remembered orientation of the stimulus that had appeared earlier at this location. The procedure in Experiment 2 was the same, except that no feedback was shown. (B) Instructions provided to the subjects in Experiment 1.

Procedure

At the start of the experiment, subjects received written instructions about the task and about how their performance would be scored (Fig. 1B). Next, they were informed about the bonus payment. For a subject in the condition with a maximum bonus of $10, the text in this screen would read "*You will receive $1 for each point you get on a trial. For example, if your highest score among the three trials is 7 points, then your actual bonus is $7. You will not receive any bonus if you get 0 points!*". Thereafter, they performed 15 practice trials that were identical to trials in the actual experiment. After finishing these trials, a multiple-choice quiz was presented with three questions to test the subject's understanding of the task and the potential bonus payment. Subjects who failed on at least one of these questions were disqualified from the experiment. The remaining subjects performed 250 trials of the delayed-estimated task with the five set sizes pseudo-randomly intermixed. To check if subjects were paying attention, we asked them at three points in the experiment to press the space bar within 4 seconds. Subjects who at least once failed to do this were presumably not paying attention and were therefore excluded from the analyses.

Results

Data from 10 subjects were excluded from the analyses because they failed to respond to at least one of the three attention-checking questions. Of the remaining 189 subjects, another 35 were excluded because they had response error distributions that did not significantly differ from a uniform distribution, as assessed by a Kolmogorov-Smirnov test with a significance level of 0.05. For the remaining 154 subjects, we computed the circular variance of the response error distribution at each set size (Fig. 2A, left). We performed a Bayesian Repeated-Measures ANOVA (JASP Team, 2018; Rouder, Morey, Speckman, & Province, 2012) on these measures, with set size as a within-subjects factor and bonus size as a between-subjects factor. The results indicated extremely strong evidence for a main effect of set size ($BF_{incl}=\infty$), but evidence *against* a main effect of bonus size $BF_{incl}=0.048$[1].

Discussion

The results of Experiment 1 showed no evidence of an effect of performance-contingent reward on VWM performance. One possible explanation of this null result is that resource allocation in VWM is insensitive to monetary reward. However, there are at least two factors in the experimental design that may have interfered with the reward manipulation. First, subjects received trial-to-trial feedback. Being constantly confronted with their own performance may have motivated them to perform as well as possible regardless of the amount of bonus they could earn. Second, since the bonus was mentioned only at the beginning of the experiment, subjects may have performed the task without having the bonus strongly on their minds. To address these potential confounds, we ran a second experiment in which subjects did not receive trial-to-trial feedback and were reminded regularly of the bonus.

[1] BF_{incl} quantifies how likely the data are under the models that include a main or interaction effect relative to how likely they are under models that do not include this effect. For example, $BF_{incl}=0.048$ for a main effect of bonus size indicates that the data are 1/0.048=~20.8 times more likely under the models that do not include this main effect compared to models that do include it.

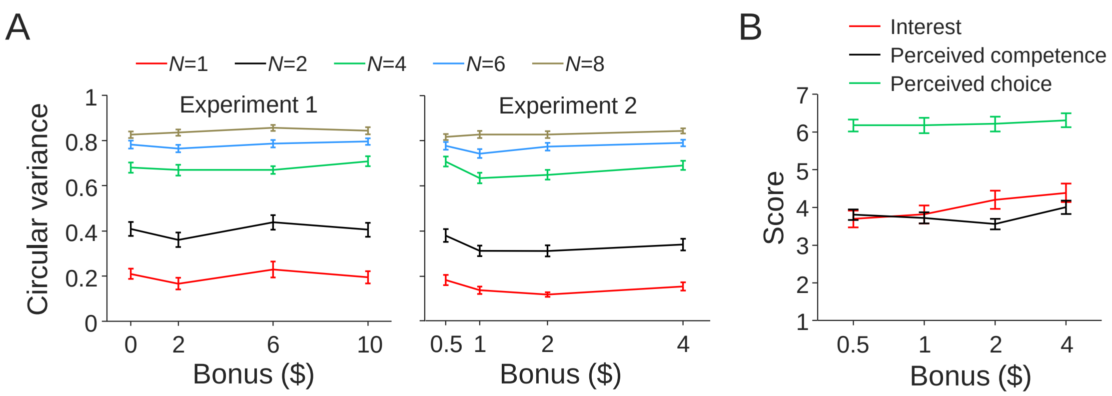

Figure 2 | Effect of bonus on VWM performance and motivation scores. (A) Subject-averaged circular variance of the estimation error distribution as a function of the amount of potential bonus in Experiments 1 (left) and 2 (right). (B) Intrinsic Motivation Inventory scores as a function of the amount of potential bonus, split by item category. Error bars indicate 1 s.e.m.

EXPERIMENT 2

Recruitment

A new cohort of subjects was recruited on the Amazon Mechanical Turk platform. The experiment was visible only to subjects who were located in the USA, had not participated in the experiment before, and had an approval rate of 95% or higher. A total of 241 subjects signed up, of whom 41 were disqualified due to failing the post-instruction quiz. The remaining 200 subjects were randomly assigned to four groups (*n*=52, 48, 50, 50) that again differed in the amount of potential bonus payment. The base payment was $5 and the potential bonus amounts were $0.50, $1, $2, and $4. The experiment was approved by the Institutional Review Board of New York University.

Stimuli and procedure

The stimuli and procedure for Experiment 2 were identical to Experiment 1, except for the following differences. First, subjects were reminded of the bonus four times in the instruction screen (compared to only once in Experiment 1) and during the task itself the following message appeared after every 50 trials: "*You have completed X% of the Experiment. Remember that you have the chance to earn a $Y bonus!*", where *X* and *Y* were determined by the number of completed trials and the amount of bonus, respectively. Second, no performance feedback was given, neither during practice nor during the actual experiment. Third, the length of the practice phase was reduced to 10 trials, but three "walk-through trials" were added at the start in which subjects were fully guided with additional written instructions. Lastly, after the experiment, subjects filled out 20 questions from the Intrinsic Motivation Inventory (McAuley, Duncan, & Tammen, 1989; Ryan, 1982) which related to their "Interest", "Perceived choice", and "Perceived competence" in the task. They rated these items on a Likert scale from 1 ("not at all true") to 7 ("very true"). The full questionnaire can be found at https://osf.io/mwz27.

Results

Data from 27 subjects were excluded because they failed to respond to one of the attention-checking questions (9 subjects) or had a response error distribution that did not significantly differ from a uniform distribution according to a Kolmogorov-Smirnov test (18 subjects). We performed the same statistical analyses as in Experiment 1 on the data from the remaining 173 subjects (Fig. 2A, right). Again, we found extremely strong evidence for a main effect of set size ($BF_{incl}=\infty$) and evidence *against* a main effect of bonus size ($BF_{incl}=0.34$). Hence, it seems unlikely that the absence of an effect in Experiment 1 was due to subjects being unaware of the potential bonus payment or due to presence of trial-to-trial feedback.

Next, we assessed whether bonus size affected the subjects' scores on the intrinsic motivation inventory questions (Fig. 2B). Using Bayesian one-way ANOVAs, we found that there was no effect in any of the three categories: $BF_{10}=0.275$ for mean "interest" scores, $BF_{10}=0.174$ for mean "perceived competence" scores, and $BF_{10}=0.034$ for mean "perceived choice" scores. Nevertheless, we noticed that there was considerable variation in the intrinsic motivation scores across subjects, especially in the "Interest" and "Perceived competence" categories (Fig. 3A). Therefore, we next tested if there was an effect of motivation scores on VWM performance. To this end, we grouped subjects from Experiment 2 into "low motivation" and "high motivation" subgroups by using a median split on each of the three categories of the Intrinsic Motivation Inventory (Fig. 3B). To examine whether scores in any of the three categories is predictive of VWM performance, we performed a repeated-measures Bayesian ANOVA with set size as within-subjects factor and motivation score ("low" and "high") as a between-subjects factor. All three tests provided evidence for the null hypothesis that there was no performance difference between subjects in the low and high motivation subgroups (Interest: $BF_{incl}=0.12$; Perceived competence: $BF_{incl}=0.21$; Perceived choice: $BF_{incl}=0.21$).

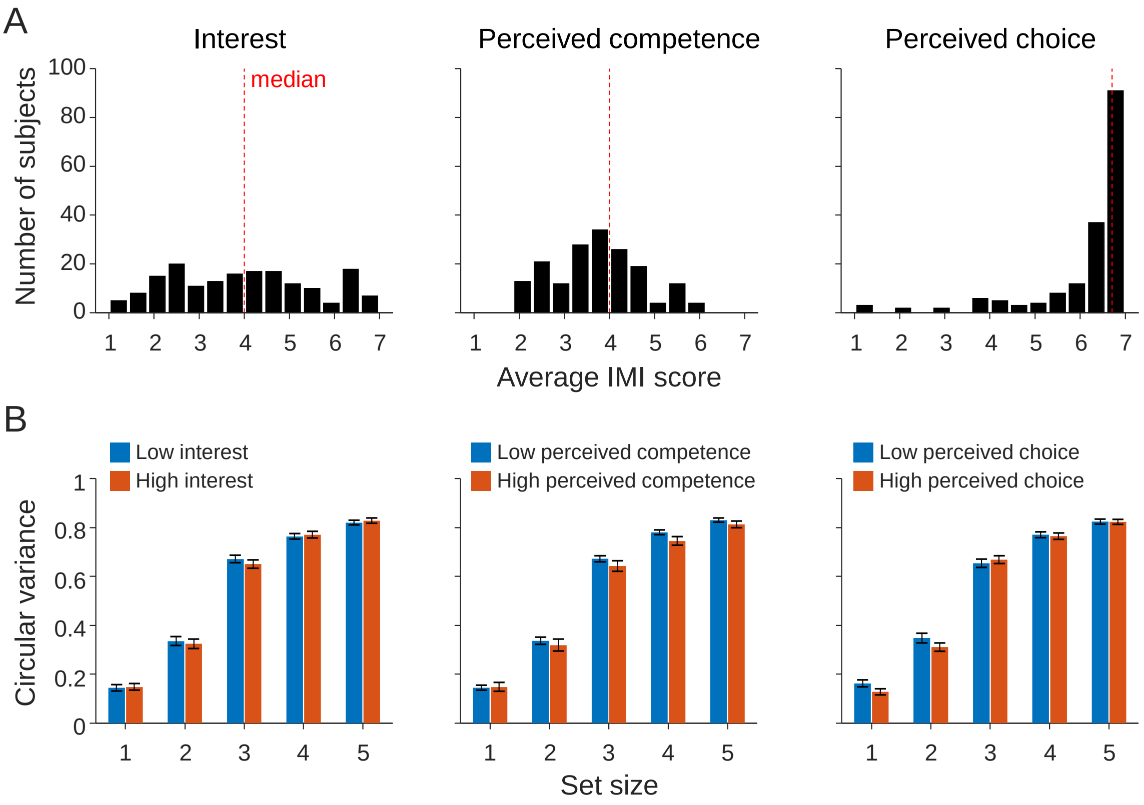

Figure 3 | Comparison of VWM performance between subjects with low and high scores on the Intrinsic Motivation Inventory (IMI). (A) Distribution of average IMI scores, split by question category. (B) Circular variance of the response error plotted separately for subjects with below-median and above-median scores on the IMI questionnaire.

Discussion

The aim of Experiment 2 was to test whether the null effect from Experiment 1 persists if we remove trial-by-trial feedback and remind subjects more often of the potential bonus. We found that this was not the case: again, there was no effect of monetary reward on VWM performance. This further strengthens the hypotheses that VWM resource allocation is independent of monetary reward. We also found that intrinsic motivation does not depend on the amount of monetary reward. This suggests that our current results are limited to the domain of *external* motivation and leave open the possibility that VWM resource allocation may be sensitive to manipulations of *intrinsic* motivation.

GENERAL DISCUSSION

In two experiments, we found no evidence that VWM resource allocation depends on performance-contingent monetary reward. We consider multiple possible explanations for this finding. First, it may be that VWM uses a fixed amount of resource, independent of the task at hand. However, this explanation contradicts previous evidence suggesting that the amount of allocated resource depends on task instructions (Bengson & Luck, 2016), set size (van den Berg & Ma, 2018), and cueing condition (Myers et al., 2018). Moreover, this kind of rigidity would stand in stark contrast to the flexibility with which VWM resource is divided among items within a trial when items have varying importance (Bays, 2014; Bays et al., 2011; Emrich et al., 2017; Gorgoraptis et al., 2011; Yoo et al., 2018; Zokaei et al., 2011). A second possible explanation for the null effects is that the bonuses may have been too small to cause an effect. We believe this to be unlikely too, especially in Experiment 1, where the bonus could increase the earnings in one of the groups by a factor 11 ($10 bonus in addition to $1 base payment). Third, subjects might not have had the bonus strongly enough on their minds when performing the task. While this explanation could be plausible in Experiment 1 – where subjects were informed about the bonus only at the very beginning of the experiment – it seems implausible in Experiment 2, where they were regularly reminded of it. Fourth, it may be that bonus manipulations are only effective when they are administered on a trial-by-trial basis, as suggested by an earlier study on the relation between task preparation and reward (Shen & Chun, 2011). Fifth, we may inadvertently have biased our subject sample to "over-performers", by only recruiting subjects who had a high approval rate on the Amazon Turk. The desire to maintain a high approval rate may have worked as a strong incentive for these subjects to perform well, regardless of the amount of performance-related bonus they could earn.

Altogether, the currently available evidence on the relation between motivation and VWM performance remains slim and mixed, which would make any strong conclusion premature. One important direction for future research would be to use a within-subject design that test effects of trial-by-trial variations in monetary reward. Another interesting direction would be to test for effects of intrinsic motivation on VWM performance, for example by "gamifying" the experiment (Hamari, Koivisto, & Sarsa, 2014). Finally, it would be worthwhile to examine whether subjects recruited on the Amazon Mechanical Turk platform are generally "over-performers", because this would have important implications for studies that examine effects of motivation on human behavior.

ACKNOWLEDGMENTS

This research was supported by grant 2018-01947 from the Swedish Research Council to R.v.d.B, training grant R90DA043849-03 to Q.Z., and grant R01EY020958-09 to W.J.M.

REFERENCES

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(10), 3632–3645. https://doi.org/10.1523/JNEUROSCI.3204-13.2014

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*. https://doi.org/10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-015-0891-7

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X01003922

Emrich, S. M., Lockhart, H. A., & Al-Aidroos, N. (2017). Attention mediates the flexible allocation of visual working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1454–1465. https://doi.org/10.1037/xhp0000398

Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quantity and quality in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1231–1240. https://doi.org/10.1037/xhp0000211

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. https://doi.org/10.1038/ncomms2237

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic Updating of Working Memory Resources for Visual Objects. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0208-11.2011

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.1109/HICSS.2014.377

JASP Team. (2018). JASP (Version 0.8.4.0) [Computer program].

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLoS Computational Biology*, *9*(2).

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413

Myers, N. E., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2018). Benefits of flexible prioritization in working memory can arise without costs. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000449

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 5975–5979. https://doi.org/10.1073/pnas.0711295105

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*.

https://doi.org/10.1037/0022-3514.43.3.450

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 277–296). Hillsdale, NJ, NJ: Erlbaum.

Shen, Y. J., & Chun, M. M. (2011). Increases in rewards promote flexible behavior. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-010-0065-7

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, *119*(4), 807–830. https://doi.org/10.1037/a0029856

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in visual working memory. *ELife*, *7*(e34963). https://doi.org/10.7554/eLife.34963

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*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. https://doi.org/10:1167/4.12.11

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*. https://doi.org/10.1038/s41598-018-34282-1

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal* *of Vision*. https://doi.org/10.1167/11.14.2

APPENDIX

Scoring functions

In both experiments, subjects received points on each trial based on the accuracy of their estimate. In Experiment 1, errors were mapped to scores through the function $s = 10 \cdot e^{-\frac{\varepsilon^2}{800}}$, where ε is the error in degree. The score was rounded to the nearest integer to obtain the number of points (Fig A1, black). In Experiment 2, a highly similar function was used (Fig. A1, red).

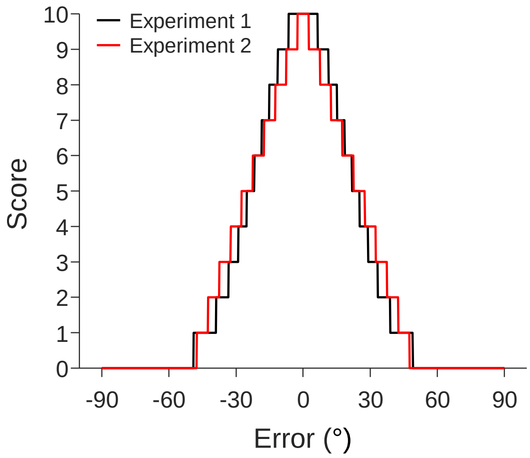

Figure A1 | Functions used to map an estimation error to a score in Experiments 1 and 2.