Does monetary reward increase visual working memory performance?

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 8, 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

ABTRACT

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 address this question by testing whether increasing monetary incentives results in better overall VWM performance. In two experiments, a total of 380 subjects performed a delayed-estimation task on the Amazon Turk platform. The base payment was $1 in Experiment 1 and $5 in Experiment 2. Additionally, subjects received a bonus payment based on their performance. The maximum bonus that a subject could receive differed between groups and ranged from $0 to $10 in Experiment 1 and from $0.50 to $4 in Experiment 2. On average, performance was better in Experiment 2 than in Experiment 1, but we found no effect of the amount of bonus on VWM performance in either experiment. These findings provide further evidence for flexibility in VWM resource allocation, but they also argue against a fully rational account of such flexibility.

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). 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 idea 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 VWM experiments in which subjects could earn a performance-based monetary bonus on top of a base payment. The base payment differed between the two experiments and the potential bonus payment varied between subject groups within each experiment. When encoding is costly, a rational observer will adjust its total amount of invest resource to the performance-related bonus, but not to the level of base payment. We found the opposite pattern: subjects performed better with a higher base payment, but were insensitive to the amount of performance-related bonus. These results provide further support for flexibility in VWM resource allocation, but they also argue against a fully rational account of this flexibility.

METHODS

Data and code availability

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

Recruitment

Subjects were recruited on the Amazon Mechanical Turk (AMT) 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. In Experiment 1, 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. In Experiment 2, 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. Both experiments were 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). In Experiment 1, feedback was provided after each trial by showing the obtained score and two lines that corresponded to the correct and responded orientations. In Experiment 2, scores were never shown to the subjects and the lines were only shown during the practice phase.

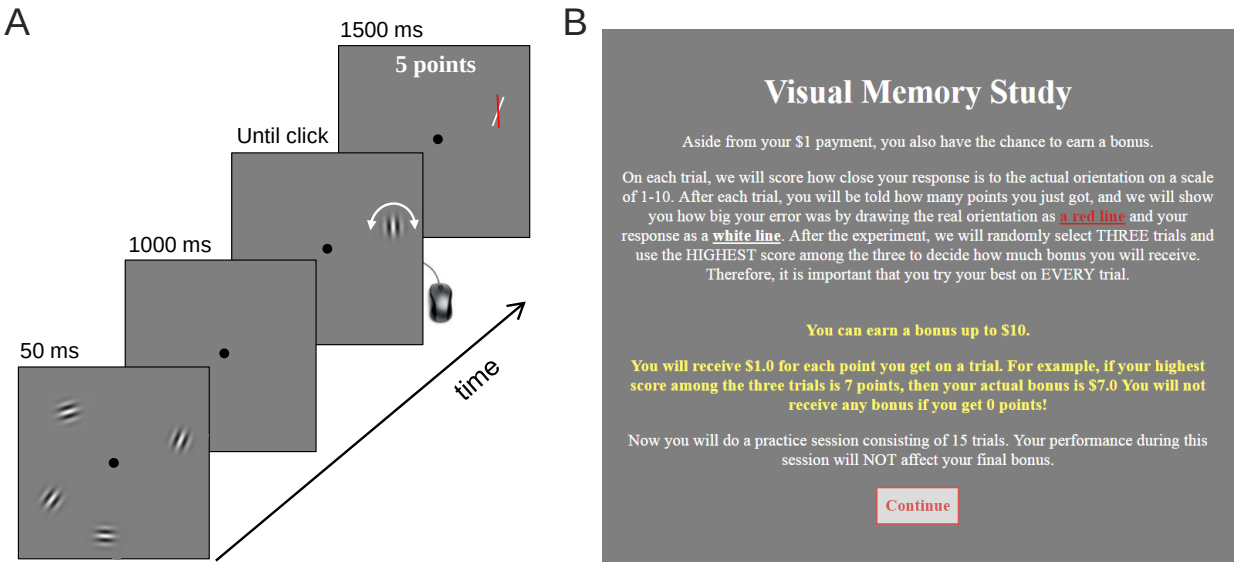

Figure 1 | Experimental procedure. (A) Illustration of a single trial in Experiment 1 (not to scale). Subjects were briefly presented with 1, 2, 4, 6, or 8 Gabor patches, which they had to keep in memory during the delay period. Thereafter, a randomly oriented Gabor patch would appear at one of the previous 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

Experiment 1. 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 was "*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.

Experiment 2. The procedure for Experiment 2 was identical to that of Experiment 1, except for the following differences. First, the length of the practice phase was reduced to 10 trials. Second, before subjects performed the practice trials, three "walk-through trials" were provided, in which they were fully guided with additional written instructions. Third, 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. Fourth, no performance feedback was given, neither during practice nor during the actual experiment. 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"). Due to a design error, the questions were all presented in the same order and category-wise rather than randomized. The full questionnaire can be found at https://osf.io/mwz27.

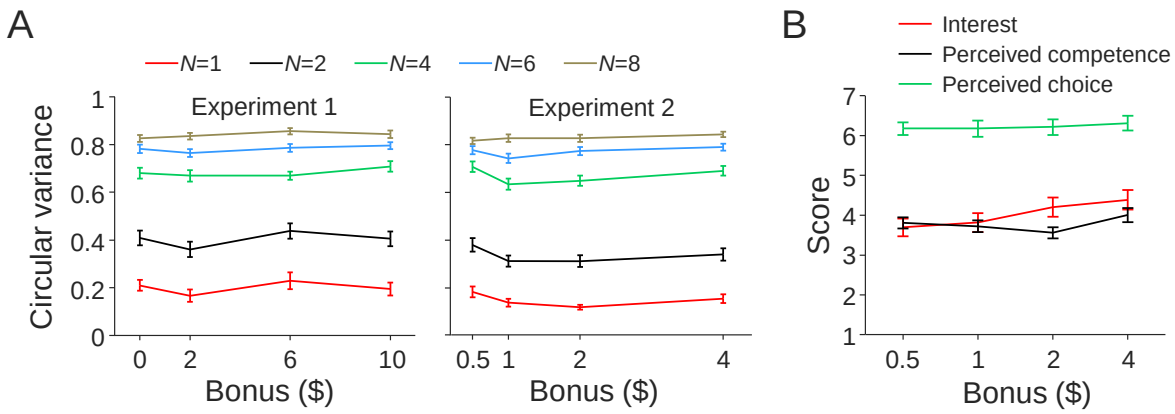

Figure 2 | Effect of bonus on VWM performance and motivation scores. (A) Subject-averaged circular variance of the estimation error distribution as a function of the amount of potential bonus in Experiments 1 (left) and 2 (right). (B) Intrinsic Motivation Inventory scores as a function of the amount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

RESULTS

Experiment 1

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 ($BF_{inclusion}=\infty$), but evidence *against* a main effect of bonus size $BF_{inclusion}=0.048$[1]. Performing the same analysis on median reaction times reveals strong evidence for a main effect of set size ($BF_{inclusion}=\infty$) and inconclusive evidence about an effect of bonus size ($BF_{inclusion}=1.34$). Hence, the amount of potential bonus for good performance did not seem to have an effect on how well subjects performed.

[1] $BF_{inclusion}$ 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, $BF_{inclusion}=0.048$ 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.

Experiment 2

In Experiment 2, we changed two several experimental factors that could have interfered with the monetary incentive manipulation. First, we removed trial-to-trial feedback, which we believe may have motivated subjects in Experiment 1 to perform well regardless of the bonus. Second, we reminded subjects about the performance-related bonus throughout the experiment (see Methods). In addition, we changed the base payment from, which allows us to examine whether there is an effect of performance-unrelated monetary reward on performance.

Data from 27 subjects were excluded because of failing to respond to one of the attention-checking questions (9 subjects) or for having 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 ($BF_{inclusion}=\infty$) and evidence *against* a main effect of bonus size ($BF_{inclusion}=0.34$). 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. A repeated-measures ANOVA on the median reaction times revealed strong evidence for a main effect of set size ($BF_{inclusion}=74.9$) and against an effect of bonus size ($BF_{inclusion}=0.16$). Finally, 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.

Combined analysis

We next performed a repeated-measures Bayesian ANOVA on the data combined across the two experiments, with experiment and bonus as between-subject factors and set size as a repeated measure. Consistent with the results above, this test revealed strong evidence in favor of an effect of set size ($BF_{inclusion}=\infty$) and against an effect of bonus ($BF_{inclusion}=0.062$). The test also provides moderate evidence in favor of an effect of experiment ($BF_{inclusion}=4.21$), which suggests that performance was influenced by the amount of base payment. Closer inspection of the data (Fig. 3) shows that the circular variance of the error distribution is consistently larger in Experiment 1 than in Experiment 2, by 26.2% (N=1), 16.4% (N=2), 3.3% (N=4), 2.2% (N=6), and 1.5% (N=8). Post-hoc t-tests revealed strong evidence that the

circular variance at set sizes 1 and 2 was higher in Experiment 1 compared to Experiment 2 (BF_{+0}=24.2 and 29.0, respectively), while there was no evidence for a difference at set sizes 4, 6, and 8 (BF_{10}=0.32, 0.31, and 0.26, respectively).

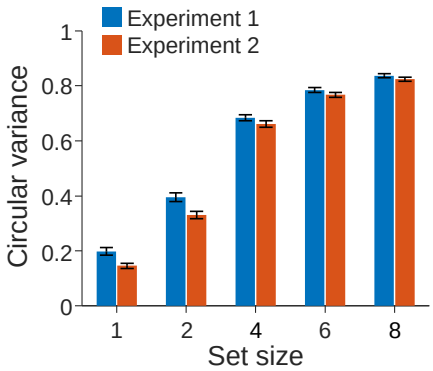

Figure 3 | Comparison of VWM performance between Experiments 1 and 2.

Effect of intrinsic motivation on performance

The findings so far suggest that there is no effect of performance-related reward on VWM performance. Moreover, the results from Experiment 2 showed that the amount of bonus did not significantly affect average intrinsic motivation scores. However, we noticed that there was considerable variation in the intrinsic motivation scores across subjects, especially in the "Interest" and "Perceived competence" categories (Fig. 4A). To test if there was an effect of motivation scores on VWM performance, we next performed an analysis in which we grouped subjects from Experiment 2 into "low motivation" and "high motivation" subgroups by using a median split (Fig. 4B). For each of the three measured categories of the Intrinsic Motivation Inventory, 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: $BF_{inclusion}$=0.12; Perceived competence: $BF_{inclusion}$=0.21; Perceived choice: $BF_{inclusion}$=0.21).

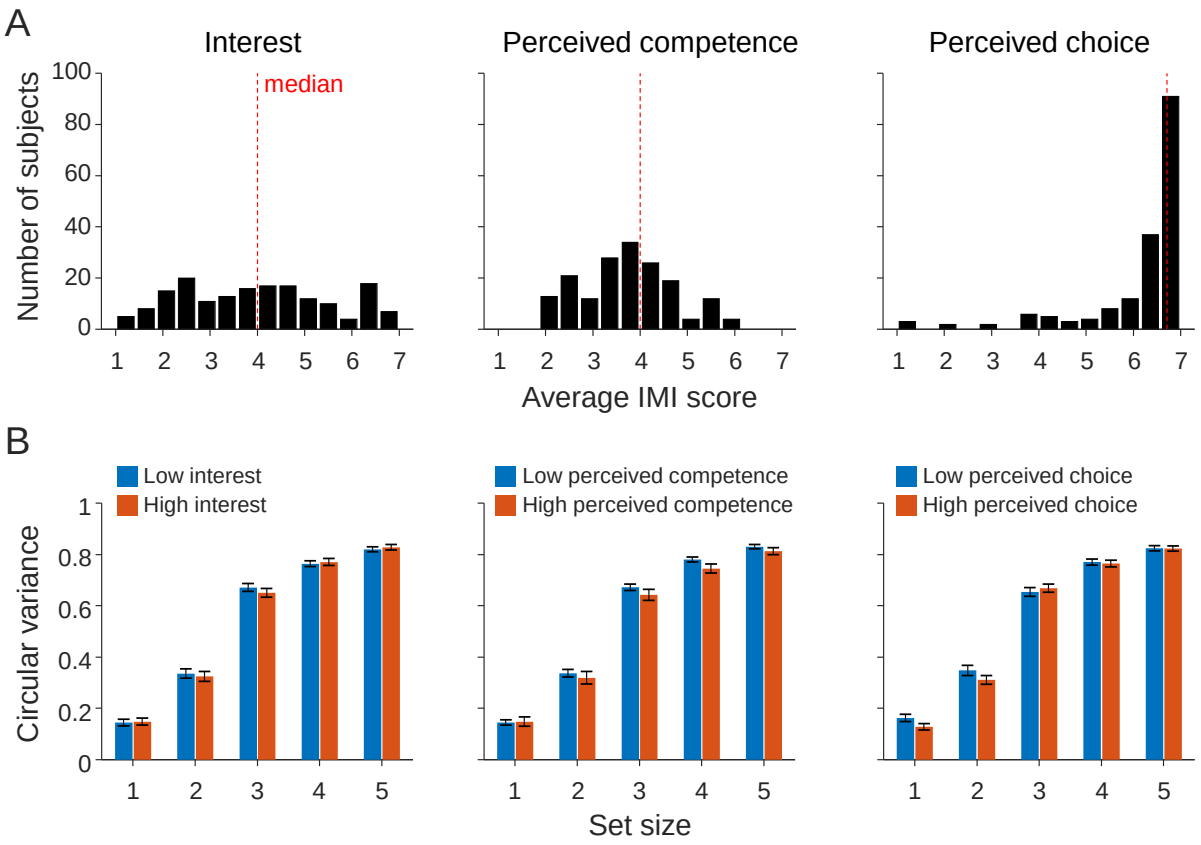

Figure 4 | Comparison of VWM performance between subjects with low and high scores on the Intrinsic Motivation Inventory (IMI). (A) Distribution of average IMI scores, split by question category. (B) Circular variance of the response error plotted separately for subjects with below-median and above-median scores on the IMI questionnaire.

DISCUSSION

We examined effects of monetary reward on VWM performance. We found that subjects who received a $5 base payment performed better than subjects who received a $1 base payment, especially at low set sizes. However, we did not find evidence for an effect of performance-related bonus on VWM performance. These results show an opposite pattern of what one would expect from a resource-rational VWM system (van den Berg & Ma, 2018), which predicts sensitivity to performance-related rewards, but not to the amount of base payment. Hence, while our findings provide further support for flexibility in VWM resource allocation, they also argue against a fully rational account of such flexibility.

From a rational perspective, it is puzzling that performance was unaffected by the amount of performance-related bonus that subjects could earn. We consider multiple possible explanations for this finding. First, it may be that subjects simply have no control over the amount of utilized VWM resource. However, this explanation is unlikely in light of the performance difference we found between Experiments 1 and 2. Moreover, it 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). Second, the bonuses may have been too small to affect a subject's incentive to perform well. 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 the bonus. Fourth, it may be that bonus manipulations are only effective when they are administered on a trial to trial basis, which would be in line with findings from 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.

Although we concluded that the amount of base payment had an effect on VWM performance, there are multiple other differences between the two experiments that could potentially explain why subjects in Experiment 2 performed better than subjects in Experiment 1. First, the bonuses in Experiment 2 were on average lower than in Experiment 1, especially when expressed relative to the base payment. However, if this had an effect, we would have expected it to go the other way around, i.e., result in poorer performance in Experiment 2. Second, subjects in Experiment 2 did not receive feedback about their performance. Again, if this had an effect, it would probably have been in the other direction. Third, subjects in Experiment 2 were reminded more frequently about the bonus. If this had made a difference, we would also have expected to see an effect of the amount of bonus on performance within Experiment 2, which we did not find. Hence, even though there were multiple differences between Experiments 1 and 2, we believe that the base payment manipulation is the most likely explanation for the observed performance difference.

A limitation of the present study is that we only investigated effects of external reward. Even if external incentives have little to no effect on VWM performance, it remains to be seen whether this also holds for factors related to intrinsic motivation. Although we included a measurement of intrinsic motivation scores, we can at present not draw strong conclusions from these data, for two reasons. First, we have these measurements for only one

of the two experiments. Second, due to a design error, the questions in the Intrinsic Motivation Inventory were all presented category-wise and in the same order to all subjects. This may have compromised independence of answers to highly similar statements (such as "I enjoyed doing this activity very much" and "This activity was fun to do"), resulting in decreased reliability of the motivation measurements. Moreover, manipulations of external rewards might have little to no effect on intrinsic motivation anyhow. Future studies could test for effects of intrinsic motivation on VWM performance more directly, for example by "gamifying" the experiment (Hamari, Koivisto, & Sarsa, 2014).

ACKNOWLEDGMENTS

This research was supported by grants 2015-00371 and 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

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).

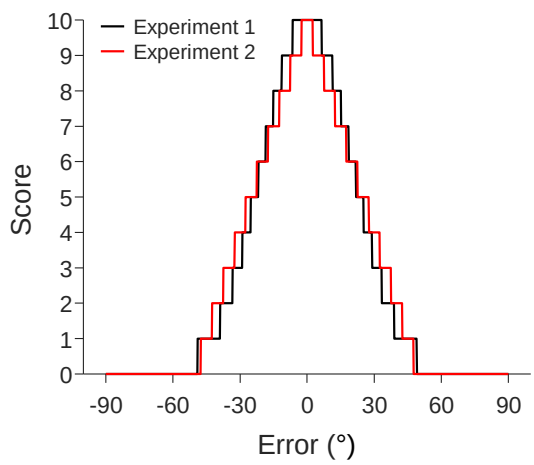

Figure A1 | Functions used to map an estimation error to a score in Experiments 1 and 2.