BRIEF REPORT

Reducing failures of working memory with performance feedback

Kirsten C. S. Adam[1] · Edward K. Vogel[1]

Abstract Fluctuations in attentional control can lead to failures of working memory (WM), in which the subject is no better than chance at reporting items from a recent display. In three experiments, we used a whole-report measure of visual WM to examine the impact of feedback on the rate of failures. In each experiment, subjects remembered an array of colored items across a blank delay, and then reported the identity of items using a whole-report procedure. In Experiment 1, we gave subjects simple feedback about the number of items they correctly identified at the end of each trial. In Experiment 2, we gave subjects additional information about the cumulative number of items correctly identified within each block. Finally, in Experiment 3, we gave subjects weighted feedback in which poor trials resulted in lost points and consistent successful performance received "streak" points. Surprisingly, simple feedback (Exp. 1) was ineffective at improving average performance or decreasing the rate of poor-performance trials. Simple cumulative feedback (Exp. 2) modestly decreased poor-performance trials (by 7 %). Weighted feedback produced the greatest benefits, decreasing the frequency of poor-performance trials by 28 % relative to baseline performance. This set of results demonstrates the usefulness of whole-report WM measures for investigating the effects of feedback on WM performance. Further, we showed that only a feedback structure that specifically discouraged lapses using negative feedback led to large reductions in WM failures.

Introduction

Failures of attention are frequent and have unintended consequences ranging in severity from variable reaction times on simple laboratory tasks to fatal car accidents in the real world (Reason, 2003; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). Given that ongoing attentional fluctuations lead to deficits in simple reaction time measures, the effect of impaired attention on demanding tasks can be even more exaggerated. Consistent with this idea, previous research has found a strong relationship between working memory (WM) capacity and propensity toward periods of mind wandering and failed executive attention (McVay & Kane, 2012). Here, we investigated the potential for feedback about task performance to reduce attentional failures during a difficult visual WM task.

Performance feedback might improve WM performance for a variety of reasons. First, subjects are relatively unaware of periods of inattention to the task at hand (Reichle, Reineberg, & Schooler, 2010; Schooler et al., 2011), but bringing subjects' attentional state into awareness allows them to re-engage (deBettencourt, Cohen, Lee, Norman, & Turk-Browne, 2015). Performance feedback should alert subjects that their current attentional state is not sufficient to perform well. In addition to alerting subjects to failures, feedback may also improve performance by increasing subjects' baseline motivation and arousal levels. Cognitive feedback can act as an extrinsic reward (Aron, 2004), and game-like visual feedback can increase subjects' intrinsic motivation

✉ Kirsten C. S. Adam
kadam1@uchicago.edu

[1]   University of Chicago, 940 E 57th St, Chicago, IL 60637, USA

Springer

(Miranda & Palmer, 2013). If subjects are relatively unmotivated in typical laboratory settings, they may underperform their true ability; providing feedback could increase task engagement and overall performance levels.

To provide informative feedback to subjects, we must first have a reliable indicator of trial-by-trial fluctuations in performance. Performance fluctuations during simple attention tasks have been extensively studied (Cohen & Maunsell, 2011; Esterman, Noonan, Rosenberg, & DeGutis, 2013; Manly, 1999; Smallwood, Riby, Heim, & Davies, 2006; Unsworth & McMillan, 2014; Weissman, Roberts, Visscher, & Woldorff, 2006), but there are few observations of performance fluctuations during complex WM tasks (Adam, Mance, Fukuda, & Vogel, 2015). Trial-by-trial performance fluctuations are difficult to measure in WM tasks because of the partial-report nature of common WM measures. For example, in a typical change-detection task, subjects are asked to remember a large number of items (e.g., six to eight) and are tested randomly on one of the items. However, capacity is extremely limited, so subjects will remember only three to four items on average. As such, even if a subject was performing quite well (e.g., three items out of six) on 100 % of trials, they would receive feedback that they were incorrect on 50 % of trials. Such unreliable feedback is unlikely to be informative to the subject.

Unlike change-detection tasks, recall tasks allow for trial-by-trial feedback about the number of correctly recalled items. We took advantage of a whole-report visual WM task to test the effects of feedback. In this task, subjects report the identity of all items in the array. Because all items are tested, performance can be calculated for every trial. Additionally, by holding set-size constant across all trials, fluctuations in memory performance can be observed without the confounding factor of intermixed difficulty from multiple set-sizes. Using this task, Adam et al. (2015) found that performance in the whole-report task was highly predictive of typical partial-report capacity measures, and that performance fluctuated strongly from trial to trial. Importantly, these results revealed that nearly all subjects have substantial numbers of WM failure trials, in which they perform no better than chance for the set of six items.

Here, we designed a series of experiments to provide different amounts of information to subjects about their trial-by-trial performance. In Experiment 1, we provided a simple form of feedback: the number correct for the trial. Simple feedback of this type is commonly used by researchers with the rationale that the feedback will increase motivation and task compliance. However, the effectiveness of such feedback is often not quantified. In Experiment 2, we added a reminder of ongoing performance by tallying the number of items correct for all trials within a block. After completing the first block of trials, subjects tried to beat their personal "high score" on subsequent blocks. We hypothesized that giving

subjects a long-term goal of improving their high score would further boost performance. Finally, in Experiment 3, we used a weighted points system instead of simple number of items correct on the trial. With this weighted points system, subjects lost points if they performed poorly and gained points if they performed consistently well. Note, however, that the subjects understood that these points were arbitrary and were not associated with any financial payout or other outcome. Unlike the other two feedback conditions, the weighted points system was designed to reinforce a particular strategy; to perform optimally, subjects needed to minimize the number of failure trials. In the other two feedback conditions, subjects could have instead attempted to maximize the number of stored items on good trials without necessarily reducing the number of poor trials. We predicted that specifically encouraging subjects to reduce the number of failure trials would maximally boost performance.

Materials and methods

Participants

All participants gave written informed consent according to procedures approved by the University of Oregon institutional review board. Participants were compensated for participation with course credit or payment (US $10/h). Forty-five subjects (21 male) participated in Experiment 1, 44 (22 male) in Experiment 2, and 56 (23 male) in Experiment 3. Subjects were excluded from analyses for non-compliance with task instructions (one in Exp. 1, one in Exp. 2 and three in Exp. 3) or for leaving the experiment early (one in Exp. 1, and one in Exp. 3).

Stimuli

Stimuli were generated in MATLAB (The MathWorks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997). Participants were seated approximately 60 cm from an 18-in CRT monitor; distances are approximate as subjects were not head-restrained. Stimuli were presented on a gray background (RGB values: 127.5 127.5 127.5), and subjects fixated a small dot (0.25° visual angle). In all experiments, colored squares (2.5°) served as memoranda. Each square could appear in one of nine colors, sampled without replacement (RGB values: Red = 255 0 0; Green = 0 255 0; Blue = 0 0 255; Yellow = 255 255 0; Magenta = 255 0 255; Cyan = 0 255 255; Orange = 255 128 0; White = 255 255 255; Black = 1 1 1) within an area extending 12.8° horizontally and 9.6° vertically from fixation. At response, a 3 × 3 grid of all nine colors appeared at the location of each remembered item. After response, a feedback screen displayed information about task performance in size 24 Arial font.

Procedures

Subjects completed a whole-report memory task in Experiments 1, 2, and 3 (Adam et al., 2015). Trial events are shown in Fig. 1a. Each trial began with a blank inter-trial interval of 1000 ms. An array containing six colored squares appeared for 250 ms, followed by a blank retention interval of 1000 ms. To respond, subjects used the mouse to click on which color appeared at each of the six locations. Subjects were required to respond to all items before the trial would proceed. Trials were blocked within experiment into two conditions, "no feedback" and "feedback." Condition order was counter-balanced across participants to prevent confounding order effects. As such, we will use the term "feedback first" to refer to subjects who received the feedback condition first and "no feedback first" to refer to subjects who received the no-feedback condition first. Subjects received 100 trials of each condition in Experiment 1 (split into four within-condition blocks), and 150 trials of each condition in Experiments 2 and 3 (split into five within-condition blocks). In between blocks, subjects received a short break (30 s in Exps. 1 and 2, 1 min in Exp. 3). In all experiments, subjects were not informed beforehand that there would be two experimental conditions. In Experiments 1 and 2, subjects completed six practice trials before beginning the first condition (two trials each of set sizes one, two, and six). In Experiment 3, subjects did not complete any separate practice trials.

In the "no feedback" condition, subjects saw a blank gray screen after responding to all items, then clicked the mouse to initiate the next trial. In the "feedback" condition, subjects saw a screen with text-based feedback about their performance. After viewing the feedback, subjects clicked the mouse to initiate the next trial. The main difference between the experiments was the content displayed on the feedback screen; an example of a typical feedback screen for each experiment is shown in Fig. 1b.

In Experiment 1, subjects received simple feedback about their performance (e.g. "3 correct"). In Experiment 2, in addition to simple feedback about the current trial, subjects received information about their cumulative "score" within a block. The cumulative score was calculated as the total number of correct objects within the block. After the first block of the feedback condition, subjects also saw their "high score", defined as the highest total performance achieved within any single block. Finally, in Experiment 3, subjects saw their current trial performance, a weighted number of "points", and a cumulative score. The cumulative score was calculated as the total number of weighted points within each block. Subjects received –2 points for 0 correct, –1 for 1 correct, 0 for 2 correct +1 for 3 correct, +2 for 4 correct and +3 for 5 or 6 correct (Table 1). In addition, "streak" bonuses were added to the weightings to emphasize consistency. If subjects earned points (three or more correct) on multiple trials in a row, a streak bonus was added to their score equaling the number of good trials. For example, if a subject got three items correct for five trials in a row, they would receive 6 points in total on that trial (+1 for three correct, + 5 for the current streak). As in Experiment 2, subjects saw their high score starting after the first block. In all three experiments, subjects received verbal instructions that the feedback information shown on the feedback screens in no way affected their payment or course credit.

Results

Experiment 1: Simple feedback

Subjects reported an average number of 2.77 (SD = .54) items correct in the no-feedback condition, and 2.82 (SD = .50) items in the feedback condition, and the difference was not significant, $t(42) = 1.90$, $P = .06$, 95 % CI [–.003, .11]. Next, we looked at the proportion of good- and poor-performance
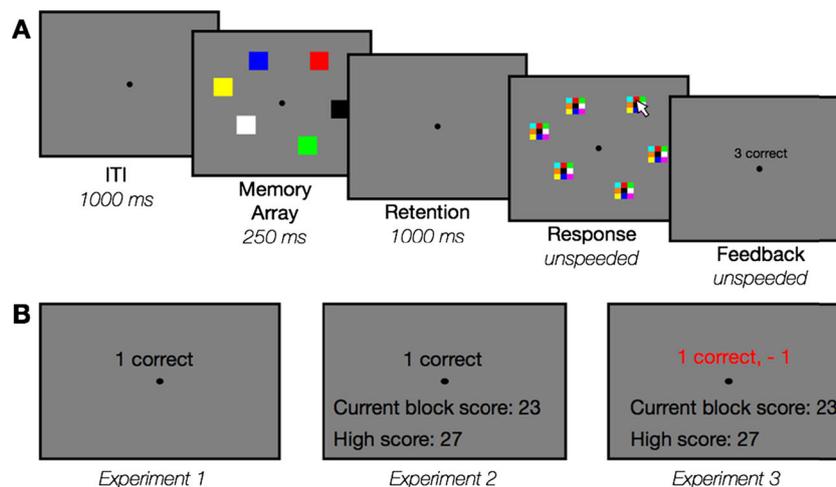

Fig. 1 **a,b** Task illustration. **a** Sequence of events for each trial. **b** Example feedback screen for each Experiment

Table 1 Points assigned for trial outcomes (number of items correct)

| Trial outcome	0	1	2	3	4	5	6
Simple feedback	0	1	2	3	4	5	6
Weighted feedback	−2	−1	0	1	2	3	3

trials (defined as above-three and below-three correct, respectively). Subjects reported more than three items correct on 25.82 % (SD = 16.88 %) of trials in the no-feedback condition versus 28.44 % (SD = 16.90 %) of trials in the feedback condition. This difference was modest but significant, indicating that subjects had slightly more good-performance trials in the feedback condition, $t(42) = 2.43$, $P = .01$, 95 % CI [.45, 4.8]. On the other hand, there was no reduction in the proportion of poor-performance trials between the no-feedback (39.95 %, SD = 19.50 %) and feedback (38.05 %, SD = 17.10 %) conditions, $t(42) = −1.52$, $P = .14$, 95 % CI [−4.4, .62]. Performance distributions are shown in Fig. 2.

We also calculated a "percent change" score for each subject to better visualize the impact of feedback on performance since there were large individual differences in the baseline rate of trial types in the no-feedback condition (Fig. 3). Percent change was calculated as the difference in the frequency of trials between the feedback and no-feedback conditions divided by the frequency in the no-feedback condition. The average percent change in good-performance trials was +20.73 % (SD = 39.80 %) relative to baseline, and the percent change in poor-performance trials was −.48 % (SD = 26.2 %). Histograms of the full range of individuals' change scores are shown in Fig. S1 for Experiment 1 and all following experiments.
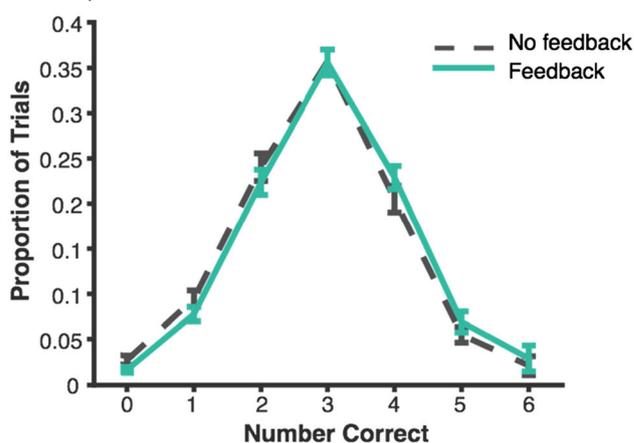
Experiment 2: Cumulative simple feedback

Subjects reported an average number of 2.87 (SD = .51) items correct in the no-feedback condition, and 3.02 (SD = .53) items in the feedback condition, and subjects performed significantly better in the feedback condition, $t(42) = 4.27$, $P < .001$, 95 % CI [.08, .22]. The proportion of good-performance trials was 27.90 % (SD = 16.10 %) in the no-feedback condition, and 32.57 % (SD = 16.72 %) in the feedback condition. This difference was significant, indicating that subjects had slightly more good-performance trials in the feedback condition, $t(42) = 3.62$, $P < .001$, 95 % CI [2.07, 7.28]. Similarly, there was a small reduction in the proportion of poor-performance trials between the no-feedback (35.69 %, SD = 15.44 %) and feedback (31.72 %, SD = 13.89 %) conditions, $t(42) = −2.94$, $P = .005$, 95 % CI [−6.7, −1.2]. Performance distributions are shown in Fig. 2. The average
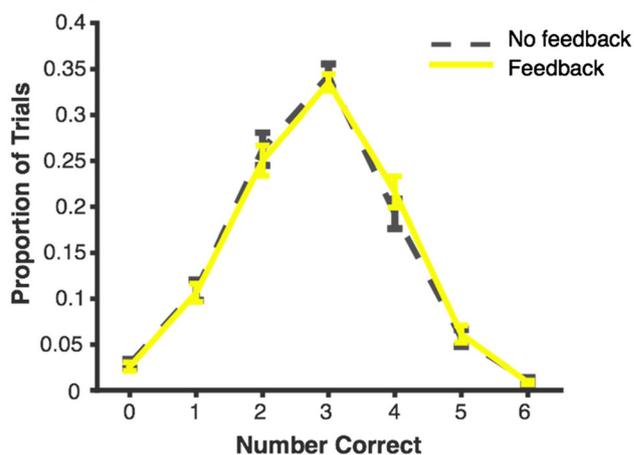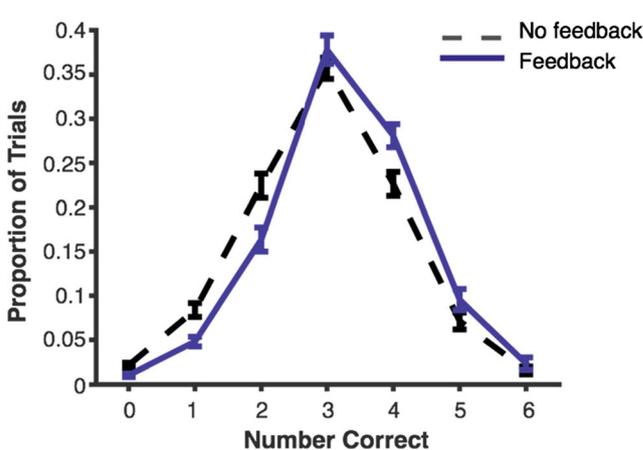

Fig. 2 Distributions of performance across experiments. *Dotted lines* Performance in the feedback condition, *solid lines* performance in the no-feedback condition in **a** Experiment 1, **b** Experiment 2, and **c** Experiment 3. All *error bars* represent standard error of the mean

percent change in good-performance trials was +30.57 % (SD = 56.27 %), and the percent change in poor-performance trials was −7.23 % (SD = 29.07 %; Fig. 3).
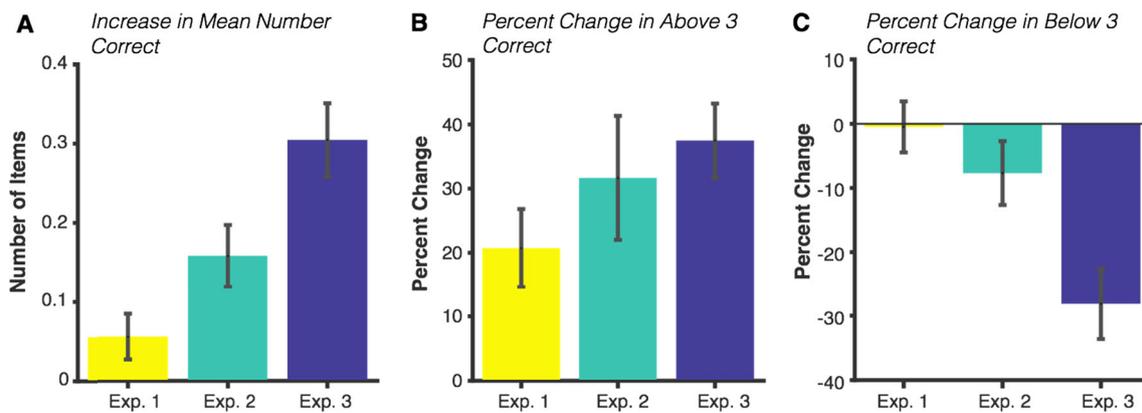
Fig. 3 **a–c** Average change in performance between the no-feedback and feedback conditions. **a** Change in the mean number of items correct. **b** Percent change in the prevalence of good-performance trials in the feedback condition relative to prevalence in the no-feedback condition. **c** Percent change in the prevalence of poor-performance trials in the feedback condition relative to prevalence in the no-feedback condition. All *error bars* represent standard error of the mean

Experiment 3: Cumulative weighted feedback

Subjects reported an average number of 2.95 (SD = .47) items correct in the no-feedback condition and 3.26 (SD = .49) items in the feedback condition, and subjects performed significantly better in the feedback condition, $t(51) = 6.70$, $P < .001$, 95 % CI [.21, .40]. The average proportion of high performance trials was 31.47 % (SD = 16.85 %) in the no-feedback condition, and 40.22 % (SD = 18.98 %) in the feedback condition. This difference was significant, indicating that subjects had more good-performance trials in the feedback condition, $t(51) = 5.92$, $P < .001$, 95 % CI [5.78 11.72]. Similarly, there was a large reduction in the proportion of poor-performance trials between the no-feedback (32.85 %, SD = 15.11 %) and feedback (22.26 %, SD = 13.37 %) conditions, $t(51) = -7.27$, $P < .001$, 95 % CI [–13.52, –7.67]. Change scores relative to baseline rates are shown in Fig. 3 (full range in Fig. S1). The average percent change in good-performance trials was +37.56 % (SD = 40.9 %), and the percent change in poor-performance trials was –27.91 % (SD = 39.1 %).

Given the general shift in subjects' performance away from poor-performance trials, we wanted to further examine the prevalence of extreme lapses (0 or 1 correct). During lapses, subjects are at chance levels of performance for this task (Adam et al., 2015), so detecting and correcting such trials would be particularly impactful in applied settings. In particular, we wanted to investigate (1) how reliably individual subjects reduced failure rates, and (2) whether the reduction in failures was consistent across all feedback blocks. Subjects had lapses during 10.49 % (SD = 7.04) of trials in the no-feedback condition, and 5.86 % (SD = 4.97) of trials in the feedback condition, and this difference is significant, t(51) = –6.19, $P < .001$, 95 % CI [–6.13, –3.13]. Thus, subjects showed on average a 33.56 % reduction (SD = 52.38) in their lapse rate (0 or 1 correct). Figure 4a shows that almost all subjects showed a lower rate of lapses during the feedback

than in the no-feedback condition, indicating that feedback interventions are effective for both high- and low-performing subjects. In addition, because the lowest-performing subjects had the highest rate of lapses, the degree of lapse reduction was negatively correlated with overall task performance, $r = -.48$, $P < .001$, 95 % CI [–.24, –.67] (Fig. 4b). However, this significant correlation was likely driven by the near-floor rates of lapses for the high capacity subjects. Previously, it was found that motivation and feedback manipulations affected high- and low-capacity subjects equally (Heitz, Schrock, Payne, & Engle, 2008). Finally, we found that performance is stable for blocks of the same condition (all Cronbach's alpha values > .78.) After a slight learning effect from block 1–2, performance was stable from blocks 2–5 with a marked shift when the condition changed in block 6 (Fig. 5).

Between experiments analyses

To compare the change in performance across experiments, we calculated change scores (feedback–no feedback) for each subject and ran a one-way ANOVA using Experiment as a between-subjects factor. First, we looked at the change in mean number of items correct and found a significant effect of Experiment, $F(2, 135) = 11.02$, $P < .001$, $\eta^2_p = .14$. Post hoc tests (Tukey's HSD) revealed that the change in mean performance was larger for Experiment 3 than for either Experiment 1 ($P < .001$) or Experiment 2 ($P = .014$). However, the change in mean performance between feedback conditions was equivalent for Experiments 1 and 2 ($P = .211$).

Similarly, we examined the change in the proportion of good- and poor-performance trials (calculated as proportion in feedback condition minus the proportion in the no-feedback condition). We found a significant effect of Experiment on the proportion of good-performance trials, $F(2,135) = 5.77$, $P = .004$, $\eta^2_p = .08$. Tukey's HSD tests revealed a significant difference between Experiment 1 and Experiment 3 ($P = .003$)
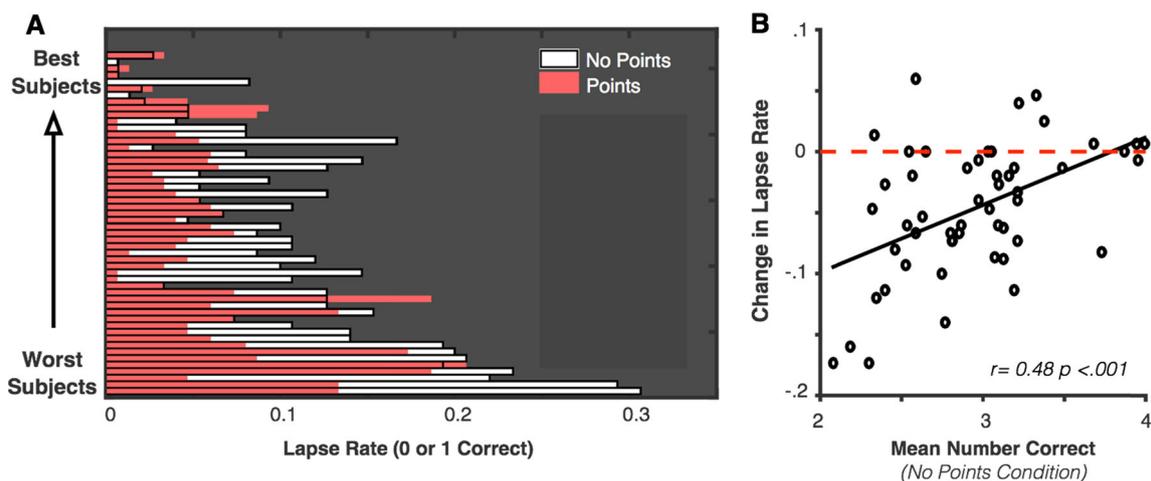
Fig. 4 a,b Change in lapse rate for individual subjects in Experiment 3. **a** Proportion of lapses (*x*-axis) is shown for individual subjects. Subjects are sorted on the *y*-axis by their overall performance in the no-feedback condition. *White bars/black outlines* Proportion of lapses in the no-feedback condition, *shaded bars* proportion of lapses in the feedback condition. Thus, if lapse rate is reduced in the feedback condition, the corresponding bar appears *white* on the *right-hand side*. If lapse rate is not reduced, then the corresponding *bar* is *entirely shaded*. Subjects in both the top and bottom of the overall performance distribution show a large reduction of lapses in the feedback condition relative to baseline lapse rates. **b** The change in lapse rate across conditions is correlated with overall task performance, likely because high-performing subjects have lapse rates near floor in both conditions

but no significant difference between Experiments 1 and 2 (*P* = .54) or between Experiments 2 and 3 (*P* = .08). Finally, we found a significant effect of Experiment on the proportion of poor-performance trials, $F(2, 135) = 11.45$, $P < .001$, η^2_p = .15. Post hoc tests (Tukey's HSD) revealed that the feedback in Experiment 3 led to a greater reduction in poor-performance trials than either Experiment 1 (*P* < .001) or Experiment 2 (*P* = .002). However, there was no difference between Experiments 1 and 2 (*P* = .56).

Discussion

We have demonstrated that a behavioral feedback manipulation can lead to global improvement in WM performance for supra-capacity arrays. Critically, however, not all forms of feedback led to the same level of improvement. These results are an important reminder to test the effects of feedback manipulations. While feedback is often assumed by researchers to always be beneficial, feedback interventions can sometimes lead to no improvement, or even to a decline in performance (Kluger & DeNisi, 1996). Moreover, the present results also suggest that estimates of maximum performance may be slightly underestimated under baseline motivational levels.

Somewhat surprisingly, providing subjects simple feedback about performance (Experiment 1) did not improve average performance, though this effect approached conventional significance (*P* = .06). The Experiment 1 feedback manipulation was similar to the feedback condition in Heitz et al. (2008), who similarly found only a small effect of feedback on
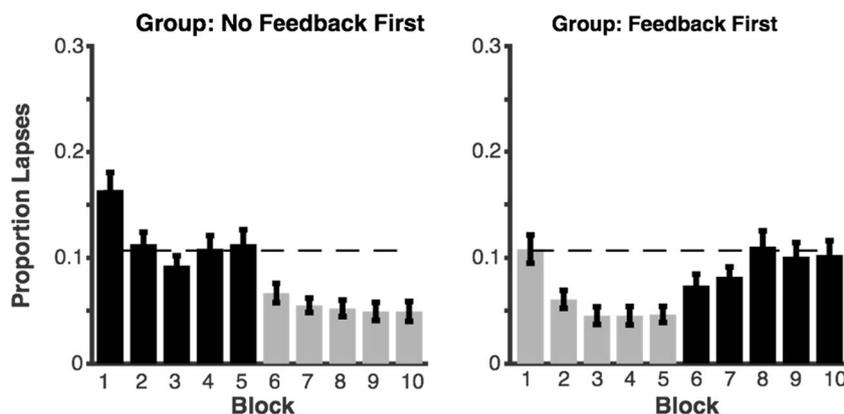

Fig. 5 Proportion of lapse trials (0 or 1 correct) over each block of Experiment 3. *Black bars* Blocks in the no-feedback condition, *gray bars* blocks in the feedback condition. *Left panel* Group of subjects who received no feedback first, *right panel* group of subjects who received feedback first. *Dotted line* Average lapse rate for blocks 2–5 during the no-feedback condition in the *left panel*

reading-span performance. In the present dataset, the marginal difference in performance was driven exclusively by an increased proportion of good-performance trials. Similarly, the increase in good-performance trials in Experiment 2 was much larger than the reduction in poor-performance trials. This asymmetry suggests that subjects attempted a suboptimal strategy of maximizing the number of items held in mind on good trials without attempting to reduce the frequency of failure trials. Overall, the persistence of lapse trials in Experiments 1 and 2 suggests that lapses of attention remain frequent and persistent in some cases where subjects are explicitly made aware of poor performance.

The weighted feedback manipulation in Experiment 3 was the most effective, reducing poor-performance trials by 27 %. Why was this manipulation so much more effective than others? Streak bonuses and the conjunction of positive and negative feedback were unique design features in Experiment 3, and they could both affect performance dramatically. Miranda and Palmer (2013) found that a visual feedback system with streaks and negative feedback increased subjects' subjective ratings of intrinsic motivation during a visual search task. Increased intrinsic motivation as a mechanism of improvement dovetails nicely with previous findings that feedback serves as an extrinsic reward (Aron et al., 2004). The conjunction of positive and negative feedback, in particular, may increase the effectiveness of feedback manipulations by engaging both pathways of the dopaminergic reward system (Frank, Seeberger, & O'Reilly, 2004). However, given the present data, we cannot say whether the addition of streak bonuses (emphasizing positive feedback) or negative feedback (punishing lapses) was most critical for performance improvement in Experiment 3. These two types of feedback are intertwined in the current design; successful streaks are perfectly anti-predictive of negative feedback. Future experiments are needed to disentangle the relative impact of each.

We would also like to emphasize the potential importance of providing an attainable performance goal in Experiment 3. Our prior work revealed that nearly all subjects are capable of accurately reporting at least three items (Adam et al., 2015). By setting a performance goal of three items, we encouraged subjects to perform consistently over a series of trials, rather than to maximize the number of items stored on individual trials. Indeed, an inappropriate performance goal could undermine the motivational benefits of feedback. If the goal was too easy (one item correct), then subjects would have incentive to underachieve their capacity. Alternatively, if the goal was too hard (six items correct), then subjects may become frustrated and similarly underperform.

Finally, our results raise some interesting questions that could be addressed by future studies. First, the observed feedback benefit dissipates shortly after the feedback is taken away (Fig. S2). However, more extensive training with feedback may help subjects learn to implement a lapse-reduction strategy without ongoing feedback. Given that behavioral feedback is relatively unobtrusive and inexpensive, there is potential for such interventions in real-world settings. Second, it will be important to disentangle the relative contributions of positive feedback, negative feedback, and performance goals to WM improvement. Finally, markers such as pupil dilation (Unsworth & Robison, 2015), and frontal theta power in EEG (Adam et al., 2015) may identify mechanisms underlying the reduction of lapses, including changes in arousal and consistency in the deployment of controlled attention.

Author contributions K.A. and E.V. designed the experiments and wrote the manuscript. K.A. collected the data and performed analyses.

Compliance with ethical standards

Conflicts of interest None

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience, 27,* 1601–1616.

Aron, A. R. (2004). Human Midbrain Sensitivity to Cognitive Feedback and Uncertainty During Classification Learning. *Journal of Neurophysiology, 92*(2), 1144–1152.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Cohen, M. R., & Maunsell, J. H. R. (2011). When attention wanders: How uncontrolled fluctuations in attention affect performance. *Journal of Neuroscience, 31*(44), 15802–15806.

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience, 18*(3), 470–475.

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex, 23*(11), 2712–2723.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306*(5703), 1940–1943.

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology, 45*(1), 119–129.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. doi:10.1037/0033-2909.119.2.254

Manly, T. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661–670.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 38(3), 525–549.

Miranda, A. T., & Palmer, E. M. (2013). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods, 46*(1), 159–172.

Reason, J. (2003). Lapses of attention in everyday life. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 515–549). London: Academic.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21*(9), 1300–1310.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*(6), 747–758.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences, 15*(7), 319–326.

Smallwood, J., Riby, L., Heim, D., & Davies, J. B. (2006). Encoding during the attentional lapse: Accuracy of encoding during the semantic sustained attention to response task. *Consciousness and Cognition, 15*(1), 218–231.

Unsworth, N., & McMillan, B. D. (2014). Fluctuations in pre-trial attentional state and their influence on goal neglect. *Consciousness and Cognition, 26*(C), 90–96.

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review, 22*, 757–865.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*(7), 971–978.