Memory & Cognition
https://doi.org/10.3758/s13421-020-01062-0

How fast can people refresh and rehearse information in working memory?

Klaus Oberauer[1] · Alessandra S. Souza[1]

© The Psychonomic Society, Inc. 2020

Abstract
Refreshing – briefly attending to an item in working memory – has been proposed as a domain-general maintenance process. According to the time-based resource-sharing (TBRS) theory, people refresh the contents of working memory sequentially at high speed. We measured the speed of refreshing by asking participants to sequentially refresh a small set of items in sync with a metronome, and to adjust the metronome to the fastest speed at which they could refresh. Refreshing speeds converged on about 0.2 s per item for several verbal and visual materials. This time was shorter than the speed of articulatory rehearsal measured with the same method, and – in contrast to rehearsal – did not depend on word length. We sought evidence for people refreshing in sync with the metronome by presenting recognition probes at unpredictable times. We expected that probes matching the just-refreshed item should be recognized faster and more accurately than probes matching other items. This was not the case. A parallel experiment with overt articulatory rehearsal showed poor synchronization of rehearsal with the metronome, suggesting by analogy that refreshing was equally out of sync. The results support the assumption that people can attend sequentially to items in working memory, and monitor this process. This refreshing process is probably faster than rehearsal, but it is unlikely to be as fast as the refreshing process assumed in the TBRS theory.

Keywords Refreshing · Overt rehearsal · Working memory · Focus of attention · Speed

Introduction

Johnson (1992) introduced the concept of refreshing as an elementary cognitive operation in the context of her MEM framework of memory. Refreshing refers to the process of briefly attending to a stimulus shortly after it has disappeared from the environment, while its representation is still in an active, available state. Barrouillet, Bernardin, and Camos (2004) reconceptualized refreshing as a maintenance process in their time-based resource-sharing (TBRS) theory of working memory. In the TBRS theory, attention is thought to rapidly cycle through the items of a memory set, refreshing them one by one to counteract decay. Hence, refreshing is assumed to work in a similar manner, and serve the same function, as articulatory rehearsal in other working memory theories (Baddeley, 1986; Schweickert & Boruff, 1986). Nevertheless, in both MEM and TBRS, refreshing is distinguished from articulatory rehearsal – the subvocal articulation of verbal working-memory contents – in that refreshing relies on domain-general central attention whereas articulatory rehearsal relies on speech production.

A computational implementation of the TBRS theory has shown that sequential refreshing counteracts decay effectively – but only when it proceeds at a very rapid rate (80 ms per item or faster) (Oberauer & Lewandowsky, 2011). At slower rates refreshing selectively boosts some memory items substantially, others not at all, and thereby creates an imbalance of memory strength that increases the risk that access to weak items is prevented by competing stronger items. A first empirical assessment of the speed of refreshing comes from the experiments of Vergauwe, Camos, and Barrouillet (2014). They asked participants to hold lists of variable length in working memory while carrying out a series of speeded decisions. To prevent rehearsal of the list, participants were further instructed to perform articulatory suppression. Mean decision times during the retention interval of the working-memory task increased by about 50 ms per item of the memory list. On the assumption that refreshing of the memory list enforces postponement of decisions, Vergauwe and colleagues inferred that refreshing proceeds at a rate of 50 ms per item. This inference is valid, however, only if people refresh the entire

✉ Klaus Oberauer
k.oberauer@psychologie.uzh.ch

[1] Department of Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland

Published online: 19 June 2020

memory list exactly once before each response to the decision task. Moreover, it is not clear whether people would spontaneously resort to refreshing when the use of rehearsal is blocked by articulatory suppression.

There are further reasons to doubt that people spontaneously refresh during the maintenance interval of a working-memory task, as several attempts to find evidence for spontaneous refreshing have failed (Vergauwe et al., 2016; Vergauwe, Langerock, & Cowan, 2018). In these studies, probes were presented at irregular intervals during the retention interval. The assumption was that if people were attending to a given item, their recognition of this item would be facilitated (as revealed by faster response times or better accuracy). The last presented item was the one recognized fastest throughout the length of the retention interval, suggesting that people kept focusing on the last presented item rather than cycling over the memory items.

Although evidence for the spontaneous use of refreshing is still lacking, people can be instructed to refresh memory items by asking them to briefly "think of" a particular item they have just encoded into working memory (Raye, Johnson, Mitchell, Greene, & Johnson, 2007; Souza, Rerko, & Oberauer, 2015). When people are instructed to refresh ("think of") a subset of items in the current memory set, their success of recalling these items exceeds that of non-refreshed items, and increases further when they are refreshed twice (Souza et al., 2015; Souza, Vergauwe, & Oberauer, 2018). Hence although it is unclear whether people use refreshing when left on their own, we can use think-of instructions to direct their attention to working memory contents, one-by-one. Here we instruct people to use refreshing (by instructing them to think of the memory items) or articulatory rehearsal during maintenance of information in working memory, thereby allowing us to test assumptions about the maximal speed at which different types of information can be refreshed and rehearsed.

This issue is of theoretical relevance for three reasons. First, working memory performance is thought to be tightly related to attention (Oberauer, 2019), and the TBRS model assumes that there is a tradeoff between how long attention is needed for refreshing and how long attention is engaged by secondary tasks (also known as distractor tasks). Accordingly, knowledge about the speed of refreshing is critical for making precise predictions about how these activities trade off to yield different levels of performance in working memory tasks. Second, refreshing is distinguished from articulatory rehearsal in that the former is a domain-general process whereas the latter is constrained to information that can be verbalized. On that account, the speed of refreshing – in contrast to that of rehearsal – should not be sensitive to parameters that affect articulation speed. Here we make a first step towards establishing such a dissociation. Third, refreshing speed has also been theoretically linked to retrieval speed, capacity limitations observed in working memory tasks, and brain oscillations (Vergauwe & Cowan, 2014). Developing a paradigm to more directly assess cyclic refreshing of memory traces may allow these assumed connections to be examined and described in further detail.

The present study

The aim of the present study was to find out how fast people can refresh items in working memory when they are instructed to do so. In a first set of studies, we asked participants to think of each item in the current memory set sequentially in a continuous cycle, and to do so in sync with a visual metronome. They were asked to adjust the speed of the metronome to the fastest speed at which they could still refresh the items. In a second set of studies, we tried to obtain behavioral evidence for the assumption that people actually refreshed in sync with the metronome, following the rationale of a study by Vergauwe and Langerock (2017): Access to the just-refreshed item should be faster, and perhaps more accurate, compared to access to other items, because the just-refreshed item is still in the focus of attention (McElree, 2006). Therefore, we interrupted the refreshing sequence at an unpredictable point with a recognition probe. We predicted that, if the probe matched the item people should have just refreshed before the probe, recognition decisions will be faster (and perhaps more accurate) than when the probe matched another item in the memory set.

In addition, we were interested in measuring the speed of articulatory rehearsal, for three reasons. First, when we investigate refreshing of verbal materials, we need to make sure to distinguish it from articulatory rehearsal. Second, there is very little evidence speaking to the speed of articulatory rehearsal. Current estimates of rehearsal speed rely on a one-page report by Landauer (1962), who asked participants to either speak aloud or "think to themselves" the contents of several lists of verbal items, and timed how long they took to go through these lists once. Landauer found that overt and silent rehearsal speeds were approximately equal. In retrospect, his instruction to "think to themselves" is very close to the instruction that Johnson, Raye, and their colleagues used to instruct refreshing. Therefore, it is not clear whether Landauer measured the speed of articulatory rehearsal or refreshing, or a mixture of both. We thought it worth revisiting the issue and comparing the speed of articulatory rehearsal to that of refreshing. Lastly, theoretical distinctions between refreshing and articulatory rehearsal rely on assumed behavioral dissociations with regard to which parameters affect attention guidance versus speech production. We sought to provide direct evidence for such dissociations.

All materials (i.e., experimental scripts and stimuli information) and data for all experiments reported here are available at the Open Science Framework at: osf.io/mcxdn; DOI: 10.17605/OSF.IO/MCXDN

Experiment 1

We asked people to hold three items in mind. In one session, participants were instructed to "think of" the items one by one in sync with a visual metronome for 10 s. In another session, they instead were to rehearse the items aloud. During and in between trials they could adjust the metronome speed to find the fastest speed at which they could "think of" or articulate the items. Memory items were colors, pictures of concrete objects, or words from one of four classes generated by crossing word length with word imageability. We predicted that word length should affect the maximal speed of articulatory rehearsal but not of refreshing. The manipulation of imageability was exploratory: We were interested in whether people understood the instruction to "think of" the words as an instruction to elaborate them, that is, to expand their representation of the word's meaning, or to form a visual image of the word's meaning. If so, then highly imageable words should be easier to "think of," and therefore could probably be refreshed faster.

Method

Participants

Twenty students of the University of Zurich took part in two 1-h sessions in exchange for partial course credit or 30 Swiss Francs (~ 30 USD). Two participants were excluded from analysis because they produced no or practically inaudible speech records in the rehearsal session, leaving a final sample of $N = 18$.

Materials

The experiment (as all the others reported here) was programmed in PsychoPy 2 (Peirce, 2007). The colors were sampled from a color circle in the CIE Lab color space (L = 70, a = 20, b = 38, radius = 60) and shown as colored disks. The pictures were sampled from 100 colored pictures of concrete objects published by Tim Brady.[1] The words were taken from the BAWL-R database (Võ et al., 2009), which contains ratings of imageability, among other variables. We selected four sets of long versus short words with high versus low imageability, holding average word frequency constant between sets (see Table 1). All words had zero orthographic neighbors so that there was no confound of word length with neighborhood size (Jalbert, Neath, & Surprenant, 2011). On each trial three elements from the relevant stimulus set were selected at random as the memory set. We chose a relatively small set size so that people could comfortably retain the items in working memory.

The visual metronome consisted of a row of three asterisks that appeared one by one at the pace of the metronome, starting in the screen center and extending to the right. With the fourth beat the metronome was reset, starting again with the first asterisk in the screen center, indicating the start of a new cycle through the memory set.

Procedure

Figure 1 shows the flow of events in a trial with words (Panel A), colors (Panel B), and pictures (Panel C). In each trial, the three memory items were presented sequentially for 0.9 s each (1.5 s for the pictures to ensure that they were encoded well, including generation of a verbal label that could be rehearsed). The words were presented centrally with an interstimulus interval of 0.1 s. The colors and pictures were presented in three equidistant locations on a virtual circle, in clock-wise order (starting at the top), with no interstimulus interval. Immediately after the last item, the visual metronome started, and continued for 10 s.

In the refreshing session, participants were instructed to "think of" each item in turn in sync with the metronome. In the rehearsal session, they were instructed to say the words aloud, or name the objects or colors aloud (overt rehearsal protocol), in sync with the metronome, and their speech was recorded. Because in a pilot experiment using only the refreshing instruction we had observed an effect of word length on the self-adjusted refreshing speed, we emphasized strongly in the refreshing instructions that they should merely think of the items and not silently speak them.

The pace of the metronome started at 0.5 s per beat, and could be adjusted during the 10-s retention interval pressing the right (faster) or the left (slower) arrow key; each key press changed the current pace by 10%. In between trials, participants also had the opportunity to adjust the metronome speed: They were asked whether they wanted to go through an adjustment phase, during which only the metronome was shown for 10 s. Alternatively, they could skip this phase and move on to the next trial directly. Participants were instructed to adjust the metronome speed to the fastest speed at which they could still "think of" (in the refreshing session) or articulate (in the rehearsal session) the items sequentially. In a pilot experiment we observed that many participants never adjusted the initial metronome speed. Therefore, we increased the speed of the metronome from each trial to the next by 10%, so that eventually the pace must become too fast, forcing participants to adjust it down.

At the end of the retention interval, memory was tested (see Fig. 1). For colors, one of the items was chosen at random and identified by a gray disk in the location of the chosen item. A color wheel with the 360 possible colors was shown at the

[1] https://bradylab.ucsd.edu/stimuli.html; we used the data set "100 Exemplar Pairs," using the first exemplar of each pair.

Table 1 Mean descriptive statistics of words used in Experiments 1 and 2

Word class	N	Number of letters	Number of syllables	Imageability	Frequency
Low-short	28	5.0	1.9	2.6	48
Low-long	60	8.8	3.0	2.7	50
High-short	33	4.7	1.7	5.2	47
High-long	49	8.6	2.9	5.2	50

same time, centered on the screen center and surrounding the positions of the three items. When participants moved the mouse from the center towards a color in the color wheel, the gray disk assumed that color, and the color was continuously updated as the mouse was moved, until the participant selected a color with a mouse click. For pictures, one item was chosen at random to be tested, and identified by a question mark in its location. At the same time, a line-up of four objects was displayed from left to right in the screen center, consisting of the three objects in the memory set and a fourth object. Participants selected the object they remembered for the given position by clicking on it with the mouse. For words, participants typed the three words in their order of presentation; each entry was prompted by a question mark displayed centrally on the screen.

There were six blocks, one for each kind of material, with 16 trials each. The adjusted pace of the metronome carried over to the next trial within each block but was reset at the beginning of each block. Order of blocks was approximately counterbalanced with a Latin square, with the constraint that not more than two blocks with words followed each other.

Analyses

We analyzed the data from this and all subsequent experiments with Bayesian linear mixed-effects models run with the *BayesFactor* package for R (Morey & Rouder, 2015). We started with the full model including all interactions, as well as individual differences (i.e., random effects of subject) on the intercept and on all main effects (i.e., random slopes).

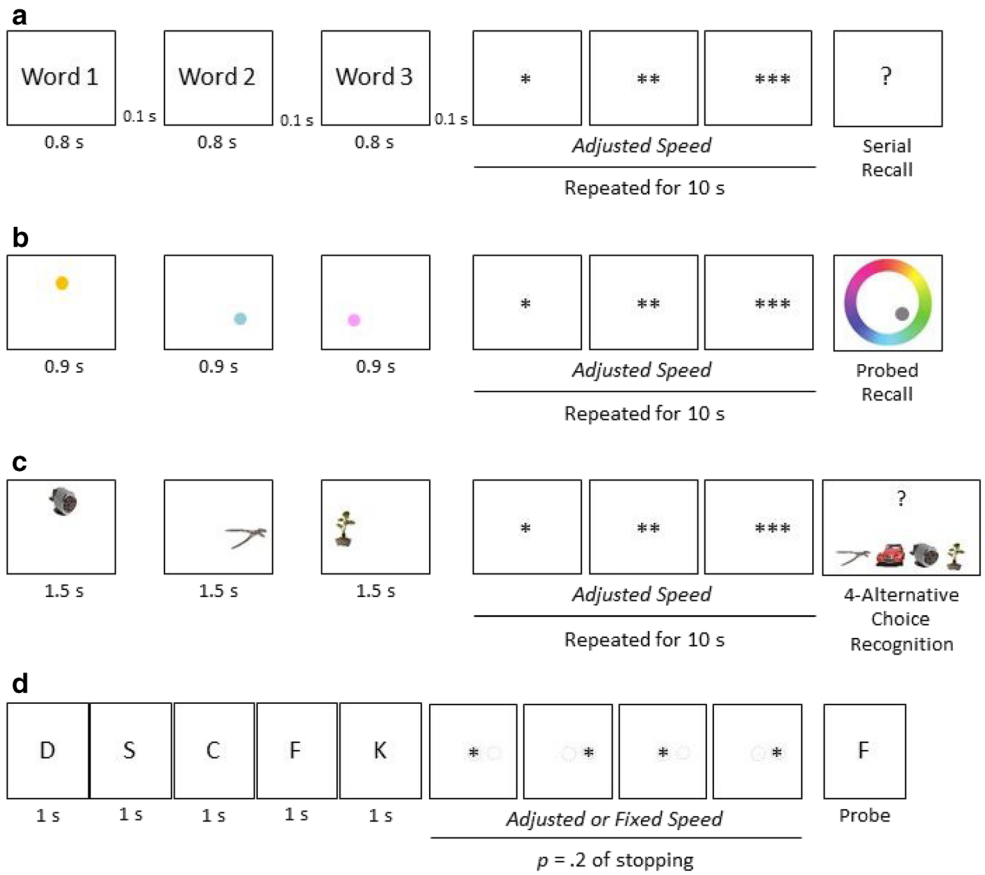

Fig. 1 Flow of events during a trial of Experiments 1 and 2 (Panels A–C) and during a trial of Experiment 3 (Panel D)

We then tested successively more constrained models by removing predictors one by one, starting with the random slopes, followed by the highest-level interaction, the lower-level interactions, and finally the main effects. At each step the more constrained model was tested against the model without the new constraint through the Bayes factor. We can think of the constrained model as the null model as it incorporates the null hypothesis with respect to the removed predictor, and the model that still includes the predictor as the alternative model. We therefore report the Bayes factor as BF_{10}, the strength of the evidence for the alternative over the null model, which reflects the evidence for an effect of the predictor in question. Whenever the Bayes factor at a comparison step favored the null model, that model was kept as the starting point for the next level of effects to be removed (e.g., when the Bayes factor favored the model without a two-way interaction, then all main effects were tested in the context of a model without that two-way interaction).

Results

By choosing a small memory load we aimed to ensure good memory accuracy, and this was successful: Recall accuracy was very good, with p(correct) > .9 for all word and picture conditions, and mean error of color reproduction at 20° and 27° in the refreshing and rehearsal conditions, respectively. Accuracy was not a dependent variable of interest and therefore we did not further analyze it.

Participants adjusted the metronome pace at least once, and mostly several times, in nearly every block. Table 2 shows the average numbers of pace adjustments participants made across the 16 trials of a block, and the percentage of blocks without any adjustments. These data show that participants complied with the instruction to adjust the speed of the metronome, as best as they could, to their maximal speed of refreshing or rehearsal.

Figure 2 shows the average metronome paces for the word conditions across the 16 trials of a block; Fig. 3 shows the paces for the colors and objects. Regardless of stimulus class, metronome pace in the refreshing condition converged on about 0.2 s per item. Paces in the rehearsal condition were more variable. In particular, long words (M over last 3 trials = .38 s) were rehearsed more slowly than short words (M = .29 s); pictures (M = .31 s) were rehearsed more slowly than colors (M = .26 s).

Word-length and imageability effects on speed

To test the prediction that word length affects the speed of rehearsal but not refreshing, we analyzed the paces, averaged over trials, with a Bayesian linear model with word length, imageability, and instruction (refreshing vs. rehearsal). Figure 4 presents the trial-averaged data. There was modest evidence against the three-way interaction, BF_{10} = 0.3. The two-way interaction of instruction with word length was strongly supported, BF_{10} ~ 600, but the interaction of instruction with imageability was not, BF_{10} = 0.5. There was also evidence against a main effect of imageability, BF_{10} = 0.2.

An analysis of the refreshing paces alone showed weak evidence against the main effect of word length, BF_{10} = 0.6, and against the main effect of imageability, BF_{10} = 0.3. By contrast, the word-length effect was strongly supported for the rehearsal paces, BF_{10} = 38447, but the effect of imageability was not, BF_{10} = 0.3. These results suggest that our instruction was successful in explaining the difference between refreshing and silent articulation, and participants did not say the words out loud to themselves in the refreshing condition. Further evidence that participants distinguished between articulatory rehearsal and refreshing comes from the observation that refreshing pace was faster than rehearsal pace, BF_{10} ~ 170.

Discussion

The results of Experiment 1 support three conclusions. First, the speed at which young adults think that they can comfortably refresh items sequentially is about 0.2 s per item. This estimate was remarkably consistent across the materials we tested (i.e., words of different lengths, colors, and pictures), as could be expected from the assumption that refreshing is a domain-general process. Moreover, the speed of refreshing

Table 2 Numbers of speed-up and slow-down adjustments of pace per block

	Experiment 1		Experiment 2		Experiment 3	Experiment 4
	Refreshing	Rehearsal	Refreshing	Rehearsal	Refreshing	Rehearsal
Mean	3.1	4.8	6.6	8.8	10.5	9.3
SD	2.9	3.1	4.8	4.8	5.7	7.3
Percent no Adjustments	27	3	7	3	0	0

Note: For each block we counted for each participant how many pace adjustments they made from one trial to the next across the 16 (Experiment 1) or 20 (Experiment 2) trials of that block; the table shows the mean and standard deviation of these counts across participants and blocks, as well as the percentage of blocks without any adjustments

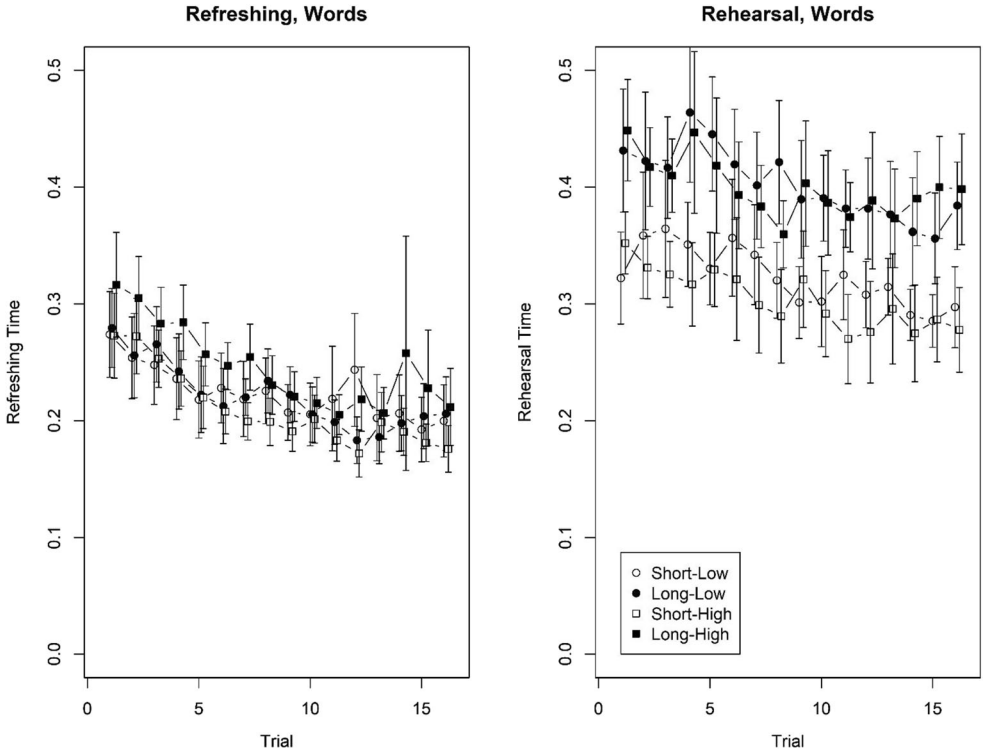

Fig. 2 Refreshing times per item to which participants adjusted the metronome while refreshing or rehearsing words in Experiment 1. Averages across participants are plotted over successive trials in each block. Error bars are 95% confidence intervals for within-subjects comparisons (Bakeman & McArthur, 1996)

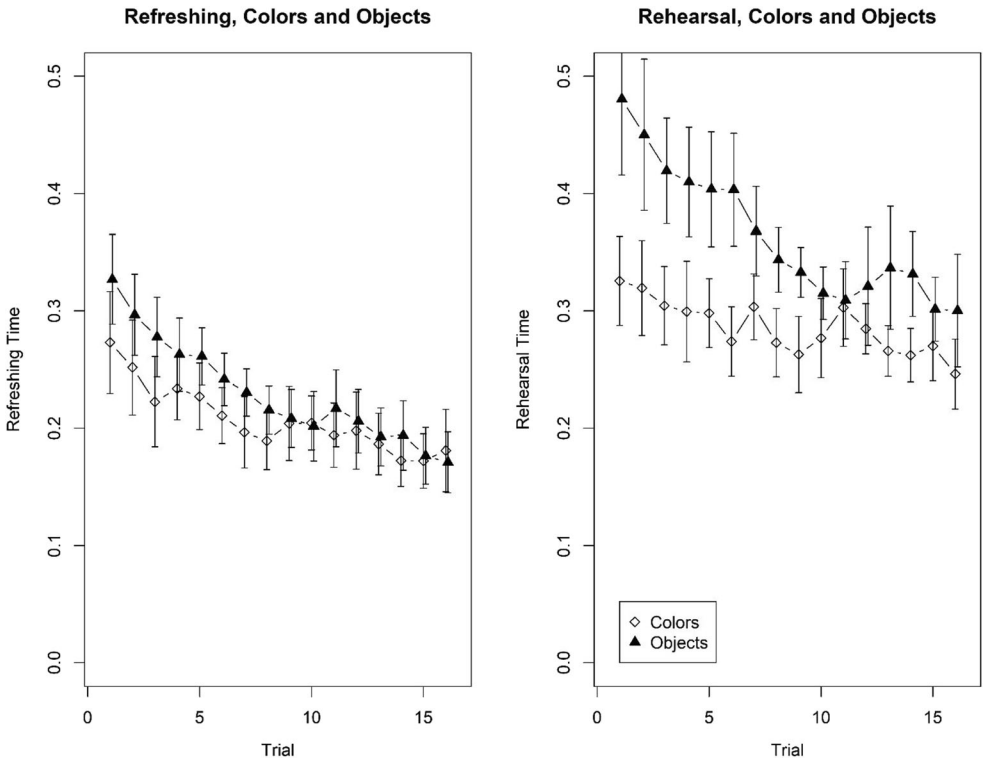

Fig. 3 Refreshing times per item to which participants adjusted the metronome while refreshing or rehearsing the labels of colors or objects in Experiment 1. Averages across participants are plotted over successive trials in each block. Error bars are 95% confidence intervals for within-subjects comparisons (Bakeman & McArthur, 1996)

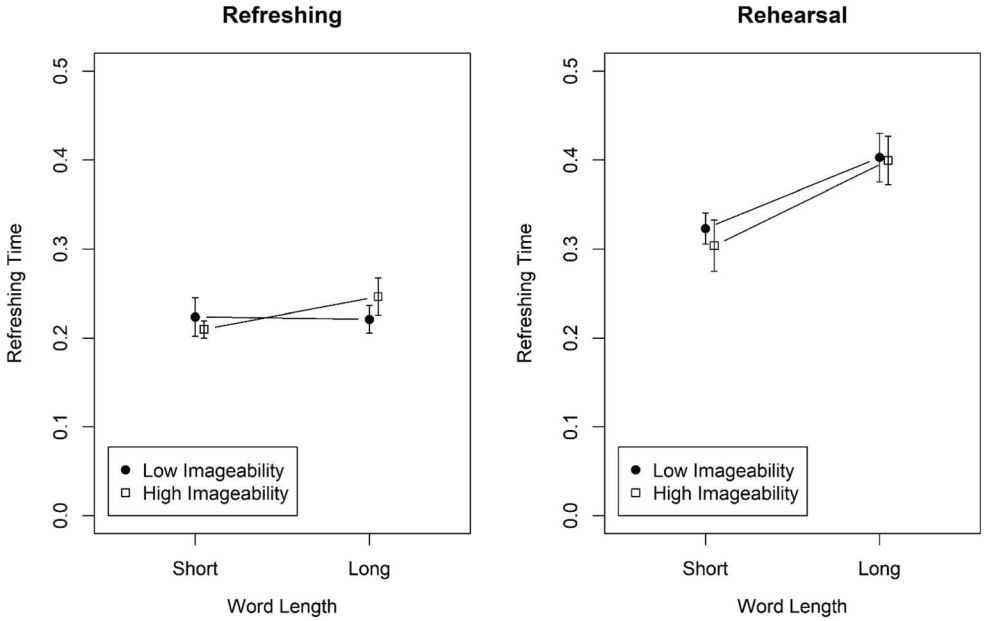

Fig. 4 Mean refreshing and rehearsal times per item for words, averaged across trials, as a function of word length and imageability in Experiment 1. Error bars are 95% confidence intervals for within-subjects comparisons

was faster than the speed at which they indicated they could articulate the items. Second, refreshing differs from articulatory rehearsal in a way that is predicted from the definitions of these processes: Rehearsal, but not refreshing, was slower for long than for short words. In addition to confirming the distinction between rehearsal and refreshing, this interaction also lends some support to the validity of people's assessments of how fast they can carry out these processes. A third conclusion is that refreshing is probably not the same as elaboration, because elaboration should be easier – and hence, faster – for highly imageable than for poorly imageable words, and we found no evidence for such a difference. This conclusion converges with other evidence for a distinction between refreshing and elaboration (Bartsch, Loaiza, Jäncke, Oberauer, & Lewis-Peacock, 2019; Bartsch, Singmann, & Oberauer, 2018).

Experiment 2

We carried out Experiment 2 with two aims in mind. First, we wanted to replicate the interaction of process instruction (refreshing vs. rehearsal) with word length. Second, we tried to obtain behavioral evidence for the assumption that people refreshed or rehearsed in sync with the metronome. To that end we drew on the work by Vergauwe and Langerock (2017), who asked participants to hold in working memory four letters displayed across the four quadrants of a 2 x 2 grid. During the retention interval, the four quadrants were highlighted one by one, in clockwise order, for 1 s each, and participants were instructed to refresh the letter they remembered for the highlighted box. After an unpredictable number of refreshing steps, a recognition probe was shown, and participants had to make a speeded decision on whether the probe matched any of the four letters of the memory set. Vergauwe and Langerock found that when the probe matched the letter that participants had just refreshed before the probe was shown, they reported the match faster than when the probe matched another letter. This response-time benefit shows that the last-refreshed item is particularly easily accessible, either because it is still in the focus of attention, or because refreshing has strengthened it to a level higher than the other items. Either way, the benefit for the last-refreshed item can be used to validate the notion that people actually refreshed the item they were instructed to refresh at any time during the retention interval. We therefore adopted this method to test whether we could find a benefit for accessing the last-refreshed item in our paradigm.

Method

Participants

Twenty-four students of the University of Zurich took part in two 1-h sessions in return for partial course credit or 30 Swiss Francs. Five participants produced incomprehensible audio records in the rehearsal session. These participants did not contribute data to the analysis of the overt-rehearsal records. One participant did not follow the instruction to rehearse the items in their order of presentation, instead rehearsing them in an erratic order; this participant was excluded (final $N = 23$).

Materials and procedure

There were three kinds of materials: Colors, short words, and long words. The colors were 36 colors selected from the color circle of Experiment 1, 20° apart from each other. The words were the same as in Experiment 1, pooling the high- and low-imageable words within each length category.

The procedure was the same as in Experiment 1, with the following exceptions: Participants were asked to refresh (in one session) or rehearse aloud (in the other session) the memory items sequentially in sync with the visual metronome throughout the retention interval, which was divided into two phases. During the first phase, lasting 5 s, the asterisks were black, and the metronome speed could be adjusted. During the second phase, the asterisks turned red, and the speed could no longer be adjusted. The second phase had an unpredictable duration: The retention interval ended with a probability of 0.2 after each metronome beat, upon which a recognition probe was shown, and participants had to make a speeded decision on whether the probe matched any element of the memory set. Participants used the mouse wheel for adjusting the metronome pace, and the two mouse buttons for the recognition decision. In this way they could use the same device for both actions, and their movements were sufficiently distinct for the two actions to minimize confusion and sequential congruency effects (i.e., biases in favor of a recognition response that is congruent with the last speed-adjustment action). Half the recognition probes matched one of the items, and of these, half matched the item that participants should have refreshed last before the probe; the other half matched one of the other two items with equal probability. In each session there were three blocks – one for each kind of material – with 24 trials each: Twelve trials with not-matching probes, six trials with probes matching the last-refreshed item, and six trials with probes matching another item, in random order.

For the plots and analyses of recognition response times (RTs) in this and the subsequent experiments, we removed all error responses as well as outliers, defined as RTs < 0.2 s or RTs exceeding a person's mean in a condition by 3 SDs.

Results and discussion

Participants adjusted the pace of the metronome in nearly every block, usually multiple times (Table 2), indicating that they followed the instruction. Figure 5 shows the paces to which the metronome was set across trials within each block. Paces from the refreshing session are on the left, and those from the rehearsal session in the middle. In addition, a research assistant listened to the audio files of overt rehearsal and pressed a button each time the participant finished rehearsing the memory set once. The last recorded time of each trial was divided by the number of rehearsed items (3 x the number of completed rehearsal cycles) to obtain an estimate of the average duration of overt rehearsal per item. These times are plotted in the right panel of Fig. 5.

Three observations can be made about this figure. First, refreshing speed again converged to about 0.2 s per item regardless of material. Second, whereas word length did not affect refreshing time (*M* short = .26 s; *M* long = .28 s), it did appear to affect rehearsal time (*M* short = .38 s; *M* long = .46 s). Statistically, however, the interaction of instruction (refreshing vs. rehearsal) and word length was not supported (BF_{10} = 0.5). Third, the speed of rehearsal as obtained from the overt-rehearsal records was considerably slower than the speed to which participants set the metronome during overt rehearsal. This observation raises some doubts about how well people can monitor their own rehearsal speed, and how well they can stay in sync with the metronome. It is worth noting that the size of the effect of word length was well matched between the self-adjusted rehearsal times and those obtained from overt rehearsal, indicating that participants were well calibrated in their estimation of the effect of word-length upon speech time.

Figure 6 presents the recognition RTs and accuracies for matching probes as a function of the deviation between the serial position of the match and the serial position last refreshed or rehearsed, as determined from the metronome beat. Positive deviations mean that the probe matched an item at a later list position than the last-refreshed position. It is obvious that there was no hint of an advantage in either speed or accuracy for a match with the last-refreshed or last-rehearsed item. We consider two reasons for this. First, participants might not have refreshed or rehearsed in sync with the metronome, so that the item they should have (according to the metronome beat) is not the one that they actually refreshed or rehearsed last before the probe. Second, with a memory set size of three we can compare probes matching the last-refreshed (or rehearsed) item only to probes matching an item at a deviation of one serial position in the refreshing/rehearsal loop. Hence, if people's refreshing or rehearsal is just one item out of sync we would no longer be able to detect any effect of what they last refreshed/rehearsed on recognition performance. Therefore, in the remaining experiments of this series we increased the memory set size to five, so that we can distinguish between absolute deviations of 0, 1, and 2. If people are slightly out of sync with the metronome, then we should be able to see a benefit for the last-refreshed item that peaks at deviation 0 and then falls off gradually with increasing absolute deviation.

Experiment 3

In Experiment 3 we asked participants to remember lists of five letters and refresh them in sync with a visual metronome.

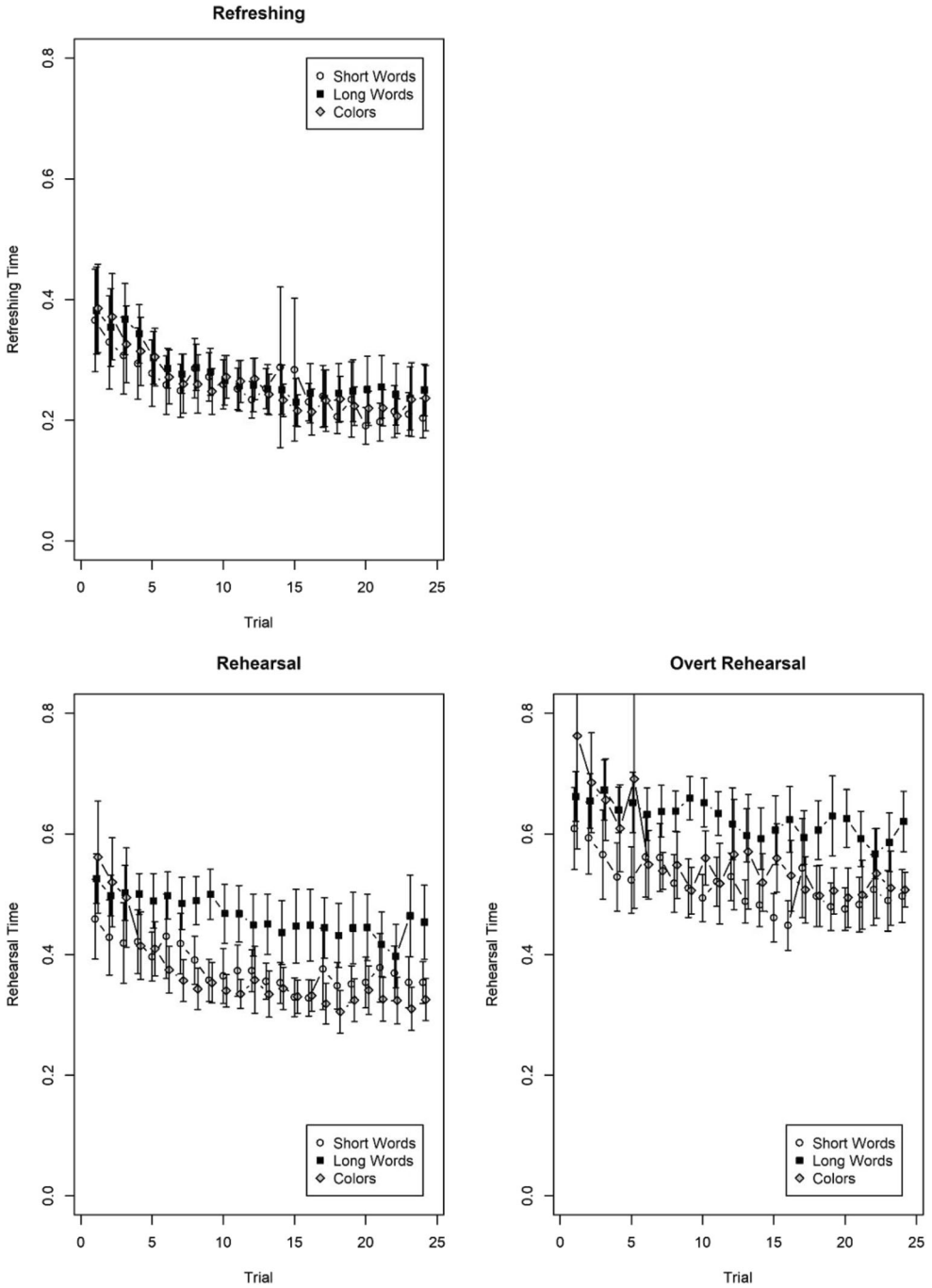

Fig. 5 Results of Experiment 2. **Top panel:** Refreshing times per item to which participants adjusted the metronome while refreshing. **Bottom panels:** Adjusted metronome time for rehearsing (left) and observed rehearsal speeds from the overt-rehearsal records (right). Averages across participants are plotted over successive trials in each block. Error bars are 95% confidence intervals for within-subjects comparisons

We turned to letters because they are easier to remember than words or colors, so with letters there is a better chance that the five-item sets do not exceed peoples' working-memory capacity.

In this experiment we changed the visual metronome to a pendulum-like alternation of asterisks in two adjacent locations. We made this change in light of several experiments in this series – reported in the Supplementary Online Material – that led us to the conclusion that people could use the visual metronome we used in the first two experiments as a cue to the list position of the item matching the probe without actually refreshing the items sequentially in sync with the metronome. This is because the row of asterisks built up cumulatively from left to right until the end of each refreshing

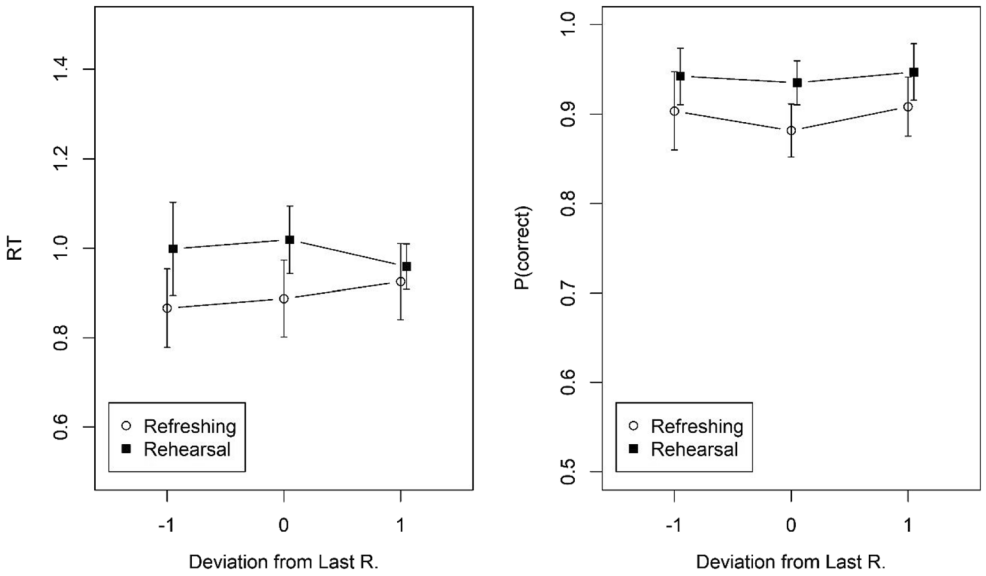

Fig. 6 Mean response times and accuracies of recognition (matching probes) in Experiment 2 as a function of the deviation between the list position that should have last been refreshed or rehearsed, and the position of the item the probed matched (positive deviations mean that the probe matched a later list position than the last-rehearsed position). Error bars are 95% confidence intervals for within-subjects comparisons

cycle, then re-started with one asterisk to mark the start of a new refreshing cycle. In this way, the number of asterisks on the screen always matched the serial position of the to-be-refreshed item. Therefore, participants did not need to follow the metronome refreshing the items continuously. Rather, they could just wait until the metronome stopped and the probe appeared, and then take the number of asterisks they had last seen on the screen as a cue to the list position of the item the probe was most likely to match to.

Briefly, the evidence showing that people used this strategy is this: In Experiments S1 and S2 we asked people to refresh five letters in sync with the cumulative metronome of Experiments 1 and 2 at five refreshing paces that we fixed (ranging from 0.1 to 1 s). As in Experiment 2, refreshing was interrupted at an unpredictable moment, and a recognition probe appeared. Different from Experiment 2, matching probes were equally likely to match any of the list items, so there was no incentive for attending to any list item more than another. We found that accuracy – but not RT – was better for probes matching the last-refreshed item than probes matching other items. Surprisingly, this was the case regardless of metronome pace – even for a pace of 0.1 s per item, which was twice as fast as the pace at which participants indicated that they could refresh.

In Experiment S3 we therefore repeated the same procedure without the visual metronome – so nothing happened during the retention interval – and without instructing participants to refresh the memory items. Simultaneously with the recognition probe a row of asterisks was displayed that looked like the last state of the metronome in the preceding experiments: The number of asterisks indicated the serial position at which the metronome would have ended, had it been displayed. In a departure from S1 and S2, a matching probe was more likely to match the item in that list position than any other list position, and participants were informed about this and encouraged to use the number of asterisks as a cue. We found the same beneficial effect on accuracy for probes matching the cued item as in Experiments S1 and S2. This shows that people can use the last state of the metronome as a cue to direct their attention to an item, and that doing so leads to the same pattern of effects as the instruction to refresh in sync with the metronome.

With the new pendulum-like metronome, the current state of the metronome could not be used as a cue to the serial position that people were supposed to refresh at any moment. Therefore, participants could consistently focus on the item they were to refresh right before the probe only if they actually refreshed continuously in sync with the metronome. Experiment 3 had two parts. In the first part participants could adjust the pace of the metronome using the mouse wheel, as in Experiment 2. In the second part, the metronome speed was fixed to one of five different paces (0.1, 0.2, 0.4, 0.7, or 1.0 s per item), and each trial ended with a recognition probe through which we tested for a last-refreshed item benefit. We varied the pace of the metronome along a broad range of values in this experiment for two reasons. First, the pace adjusted by participants may reflect a conservative (and comfortable) pace at which they can refresh items sequentially, but not their maximal speed. Participants may be able to refresh faster if they have to, but this may come at some small costs which they are trying to avoid by setting the metronome at a slower rate. Imposing a pace faster than what participants would self-

select will allow us to assess this possibility. Second, our manipulation of refreshing paces allows us to test the effect of refreshing on recognition performance across a broad range, from a value close to the theoretically assumed fast speed of refreshing (i.e., 100 ms), to values close to what participants self-selected in the preceding experiments (i.e., 200 ms), to slow and conservative paces used in previous research using cues to guide refreshing (i.e., 1,000 ms; see Vergauwe & Langerock, 2017). This will allow us to compare our results to values proposed by current theories of working memory and also to previous research, thereby indicating which refreshing speed is most beneficial to memory performance.

Method

Participants

Thirty students of the University of Zurich took part in a single 1-h session for partial course credit or 15 Swiss Francs as compensation.

Materials and procedure

On each trial five randomly selected consonants were presented at a rate of one per second in the center of the screen. This was followed by a series of asterisks presented one by one at the current metronome pace, alternating between the screen center and a position slightly to the right of center. Participants were instructed to refresh the letters in sync with the metronome, in their order of presentation.

In the adjustment part of the experiment, the retention interval was 10 s, during which participants could adjust the metronome speed with the mouse wheel as in Experiment 2. They could adjust the metronome speed in between trials as well if they chose to do so. The adjustment part ended when the adjusted metronome speed reached convergence. The criterion for convergence was that there were at least four reversals of adjusted speed (i.e., the direction in which the speed was adjusted from one trial to the next changed at least four time), and the slope of the adjusted speeds over the last eight trials was less than 5% of the mean speed. If that criterion was not met before, the adjustment phase ended after 60 trials.

The fixed-pace part consisted of five blocks of 30 trials each, within which the metronome pace was fixed to one of the values 0.1, 0.2, 0.4, 0.7, or 1.0 s per item. The order of blocks was counterbalanced across participants with a Latin square, with the constraint that paces adjacent in rank order were not run in successive blocks. Each trial's retention interval ended after an unpredictable number of metronome beats, with a constant probability of 0.2 of ending after each beat. The last metronome asterisk was followed by a centrally presented recognition probe, displayed at the time at which the next asterisk would have been shown, so that the probe onset continued the beat. Participants were instructed to decide as quickly and accurately as possible whether the probe matched any of the list letters. Half the probes matched a list item, with an equal probability of matching each item.

At the end of the experiment, participants were given another six trials – one for each pace, including the five fixed paces and the pace they had adjusted themselves – with ten beats of the metronome during the retention interval. After that they were asked to rate their confidence that they could refresh the letters at the given pace. They indicated their confidence by a mouse click on a continuous scale, displayed as a horizontal line with the end points marked as "0%: completely sure that no" and "100%: completely sure that yes," and intermediate positions labeled by the corresponding percentages in steps of 10. The responses were scaled to a range from 0 to 1.

Results and discussion

Figure 7 shows the mean RTs and accuracies for recognition of matching probes as a function of the deviation between the list position of the matching probe from the list position participants should have last refreshed in sync with the metronome. There was no hint of a benefit for the last-refreshed item. This was confirmed by Bayesian linear mixed-effects models with pace and deviation as predictors, and either mean log(RT) or proportion of correct match responses as dependent variable. In addition, we tested for a quadratic trend of deviation at each pace level separately. Table 3 shows the Bayes factors in favor of all fixed effects. The only effect supported by strong evidence was the main effect of pace on RT, which reflected slower RTs at slower paces. Apparently, participants adapted their response speed to some extent to the speed of the metronome.

In the block in which the metronome pace could be adjusted, all participants made several adjustments (see Table 2), showing that they engaged with the task to match the metronome pace to their refreshing speed also with the pendulum-style metronome. Figure 8 shows the rated confidence of being able to refresh the letters in sync with the six paces. Confidence was highly variable at paces 0.1 and 0.2 s per item, and converged on a high level for slower paces. The figure also shows – in red – the distribution of paces to which participants adjusted the metronome speed (M = 0.34 s; SD = 0.16 s). This speed was considerably slower than the adjusted metronome speeds in Experiments 1 and 2, and also compared to Experiment S1, which differed from Experiment 3 only in the way the metronome was presented (building up cumulatively vs. pendulum-like). This difference is a hint that the pendulum-like metronome made it harder to refresh items in sync with it. Some participants were no longer very confident at the end of the experiment that they could refresh at the rate

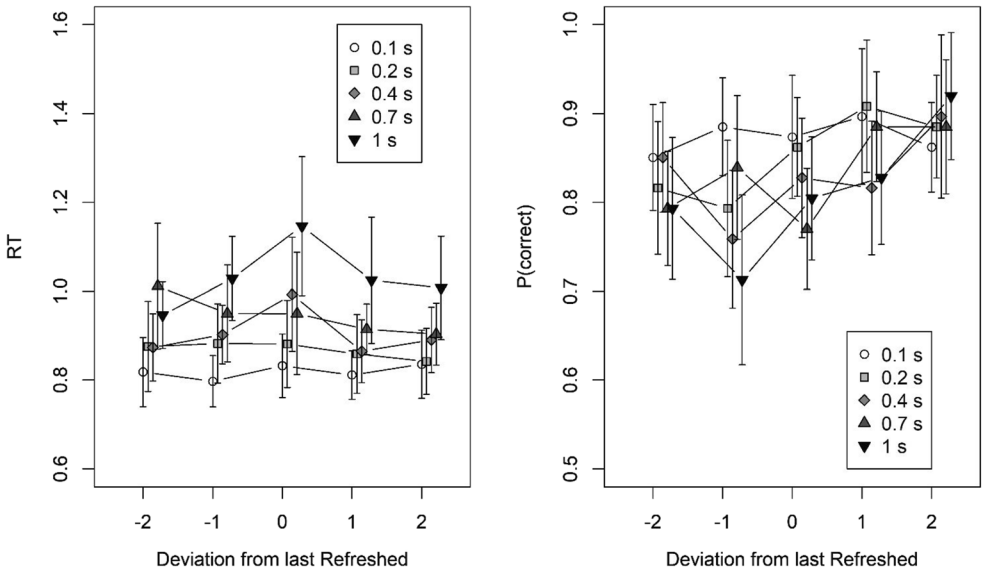

Fig. 7 Mean response times and accuracies of recognition (matching probes) in Experiment 3 as a function of the deviation between the list position that should have last been refreshed, and the position of the item the probed matched. Positive deviations mean that the probe matched a later list position than the last-refreshed position. The points legend denotes the pace condition. Error bars are 95% confidence intervals for within-subjects comparisons

they adjusted the metronome to in the first phase of the experiment (this was also the case in Experiment S1). The adjusted metronome speeds might be an overestimation of the speed at which people can actually refresh the items comfortably.

Why was there no benefit for recognition of a probe matching the last-refreshed item? One possibility is that participants did not follow the instruction to refresh the items in sync with the metronome. Another possibility is that they tried to comply with the instruction but were unable to remain in sync with the metronome, even for a few beats (on average 4.9), and even at the slow pace of 1 s per item. The overt-rehearsal data from Experiment 2 suggested that this might be the case. We therefore carried out Experiment 4, repeating the design and procedure of Experiment 3 but replacing refreshing with overt rehearsal as this is the observable behavior most similar to sequential refreshing.

Experiment 4

Method

Participants were 32 students of the University of Zurich. Experiment 4 was identical to Experiment 3 except that the instruction to refresh the items was replaced by the instruction to overtly rehearse them. Because we thought that speaking letters at a pace of 0.1 s per item was not feasible, we dropped that condition, leaving four fixed paces (0.2, 0.4, 0.7, and 1.0 s per item).

Results

The confidence ratings for being able to rehearse the letters at six different paces are shown in Fig. 9. Similar to refreshing in Experiment 3, participants were reasonably confident to be

Table 3 Bayes factors of fixed effects on recognition response times and accuracies in Experiments 3 and 4

DV	Pace	Deviation	Pace x deviation	Quadratic trend				
				0.1 s	0.2 s	0.4 s	0.7 s	1.0 s
Experiment 3								
RT	1608	0.02	0.002	0.37	0.26	1.05	0.32	0.94
Accuracy	0.06	0.88	0.06	0.26	0.36	0.48	0.33	2.09
Experiment 4								
RT	577	1.4	0.01		0.30	1.00	0.24	0.41
Accuracy	0.02	1.4	0.01		0.29	0.46	0.60	0.33

DV dependent variable, *RT* reaction times

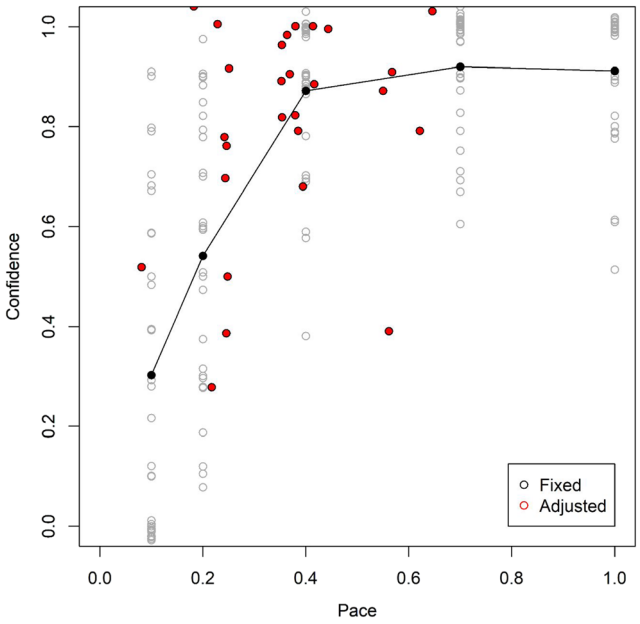

Fig. 8 Confidence ratings of how well participants could refresh in sync with the metronome at a given pace in Experiment 3. Filled black dots are means across participants for each fixed pace; unfilled dots are individual participants' ratings. Red dots are individual participants' ratings at their adjusted pace

able to rehearse the letters at a pace of 0.4 s per item or slower. Their adjusted metronome speed, reflected by the red dots (M = 0.29, SD = 0.16) was even a bit faster than in Experiment 3.

The audio records of overt rehearsal were timed in the same way as for Experiment 2: A research assistant pressed a button each time the participant finished rehearsing the list once (i.e., the time of one complete rehearsal cycle). On a substantial proportion of trials participants rehearsed fewer than five letters, and, occasionally, more than five letters; in these cases, the time at which people started repeating whichever set they rehearsed was timed as the end of one rehearsal cycle. We estimated the average speed of rehearsal by dividing the retention interval of each trial by the number of rehearsed letters. Figure 10 plots the distribution of rehearsal speeds of individual trials estimated from the overt-rehearsal records as a function of the metronome speed. It is clear that there was substantial variability in how fast people rehearsed at each metronome speed. At the faster metronome speeds, their observed average rehearsal speed was slower than the speed of the metronome, replicating the finding of Experiment 2.

If people dropped some letters from rehearsal, they could rehearse slower than the metronome but still remain aligned with the list position they should rehearse at any time. If so, the observed times at which they finished each rehearsal cycle should correspond to the time at which a rehearsal cycle ended according to the metronome beat (i.e., after every 5 x pace interval). This was not the case, as shown in Fig. 11. For all rehearsal speeds but the slowest, the observed cycle finishing times systematically lagged behind the normative ones given by the metronome.

As a consequence, on a substantial proportion of trials the letter that people rehearsed last before the recognition probe did not match the one they were supposed to rehearse last according to the metronome (see Fig. 12). Therefore, we asked whether there was a benefit for recognition performance if the recognition probe matched the letter people actually had rehearsed last.

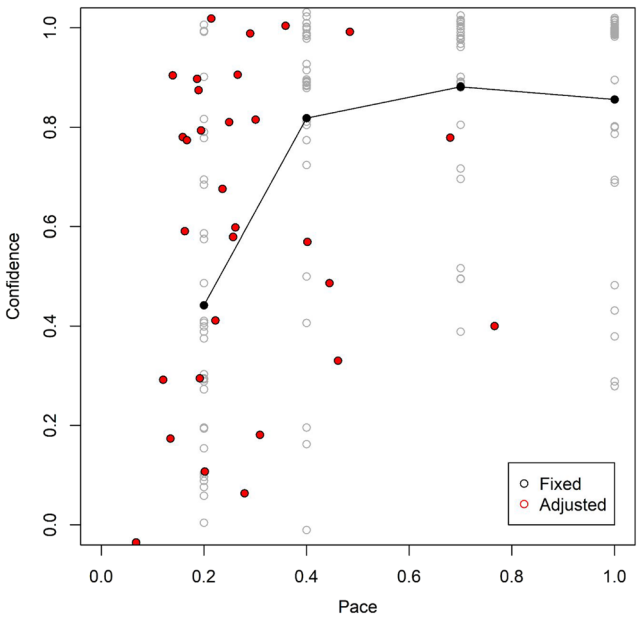

Fig. 9 Confidence ratings of how well participants could rehearse in sync with the metronome at a given pace in Experiment 4. Filled black dots are means across participants for each fixed pace; unfilled dots are individual participants' ratings. Red dots are individual participants' ratings at their adjusted pace

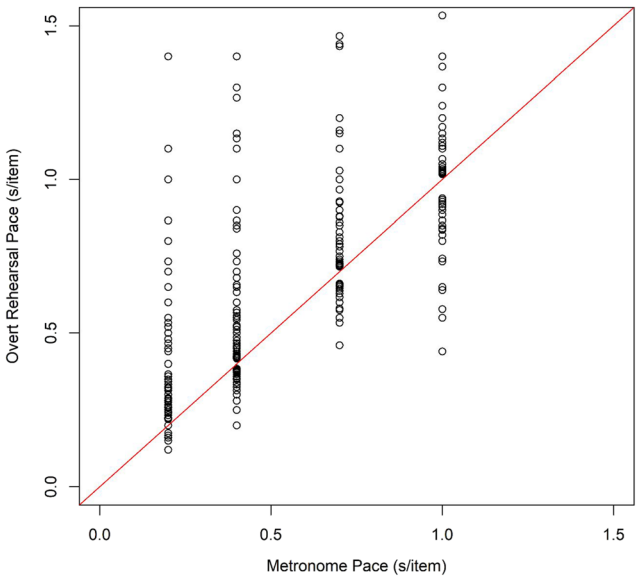

Fig. 10 Pace of overt rehearsal as a function of metronome pace in Experiment 4. Each dot is the rehearsal pace estimated from the overt-rehearsal record of one trial

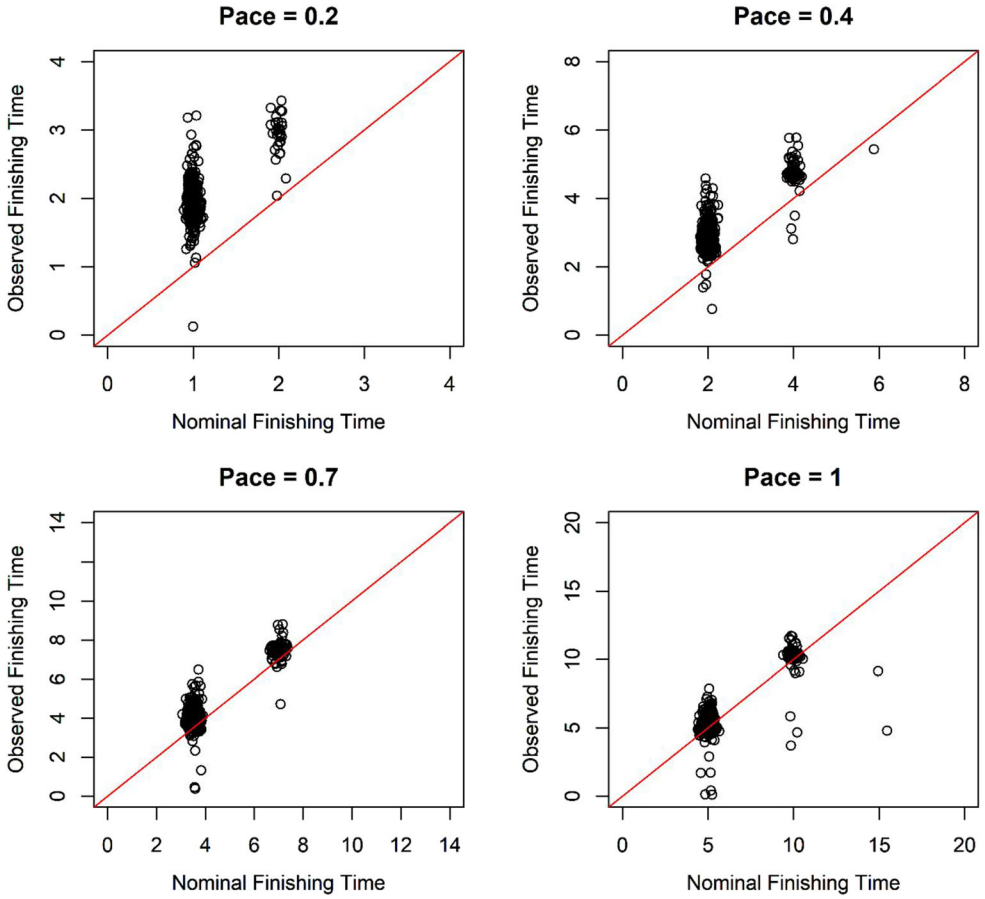

Fig. 11 Finishing times of rehearsal cycles obtained from the overt-rehearsal records as a function of the normative finishing time given by the metronome pace. Each dot is the finishing time of one rehearsal cycle in one trial

Figure 13 shows the mean RTs and proportions correct on matching recognition probes. There was no benefit for probes matching the last-rehearsed item (deviation 0) in either RT or accuracy. There appears to be a trend for slower responses and lower accuracies at deviation -1 (i.e., the probe matched the list item right before the last rehearsed item). The statistical evidence for an effect of deviation on RT (BF = 1.4) or on accuracy (BF = 1.4) was, however, weak at best (see Table 3).

General discussion

How fast can people refresh items in working memory sequentially? The self-adjusted metronome speeds mostly converged on an estimate of about 0.2 s per item regardless of the kind of material. This estimate, however, might be tied to the cumulative way of presenting the visual metronome that we used in Experiments 1 and 2 (and in Experiment S1), which gives a visual cue to the serial position to be refreshed at each moment. With the pendulum-like metronome we used in Experiment 3, the adjusted metronome speed during refreshing was slower.

These estimates are in good agreement with estimates from several experimental paradigms used to assess the time it takes to focus attention to a new item in working memory: The time cost of switching from one item to another in a memory set is about 0.3–0.7 s, depending on memory set size, for verbal materials (Garavan, 1998; Oberauer, 2003; Oberauer, Wendland, & Kliegl, 2003), and between 0.1 and 0.2 s for visual-spatial materials (Gehring, Bryck, Jonides, Albin, & Badre, 2003; Hedge & Leonards, 2013; Hedge, Oberauer, & Leonards, 2015). In experiments using retro-cues during the retention interval to direct attention to one item in a visual array, it takes about 0.3 s between cue and test for the cue to become fully effective (Souza, Rerko, & Oberauer, 2016; Tanoue & Berryhill, 2012).

Our estimate of refreshing speed is much slower than the speed of refreshing – 50 ms per item – assumed by Vergauwe and Cowan (2014) and Vergauwe et al. (2014). The evidence underlying this estimate is the observation that RTs of a secondary task during the retention interval of a short-term memory test were slowed by about 50 ms per memory item. One way to explain this finding on the assumption that people refresh at a rate of 200 ms per item is that not every item is refreshed before every response to the secondary task.

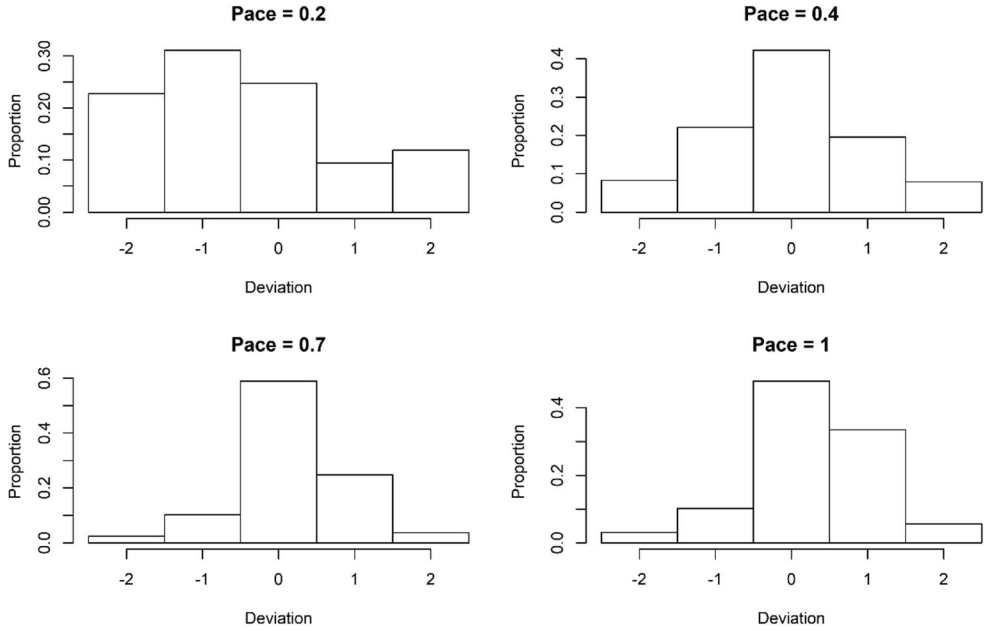

Fig. 12 Distributions of deviations between normative and actual last-rehearsed position in Experiment 4. Positive deviations indicate that the actually last-rehearsed item was in a position later in the list than the one participants should have rehearsed last according to the metronome

Alternatively, people might not refresh at all during the retention interval – as suggested by a study revisiting the experimental paradigm of Vergauwe et al. (2014) (Thalmann et al., 2019). The slowing of secondary-task RTs might have an entirely different cause.

Our estimate of refreshing speed is also much slower than the slowest speed – 80 ms per item – for which a computational implementation of the TBRS theory could reproduce the data of typical complex-span experiments (Oberauer & Lewandowsky, 2011). One way to reconcile the model with a slower refreshing rate is to assume that refreshing is not sequential; rather several items can be refreshed at the same time (Lemaire, Pageot, Plancher, & Portrat, 2018). We note, however, that this model would have to assume that parallel refreshing is slower when memory set size is larger (probably as a linear function of set size, thereby mimicking serial refreshing) to explain the data of Vergauwe et al. (2014). Empirical evidence for the conjecture that refreshing (and

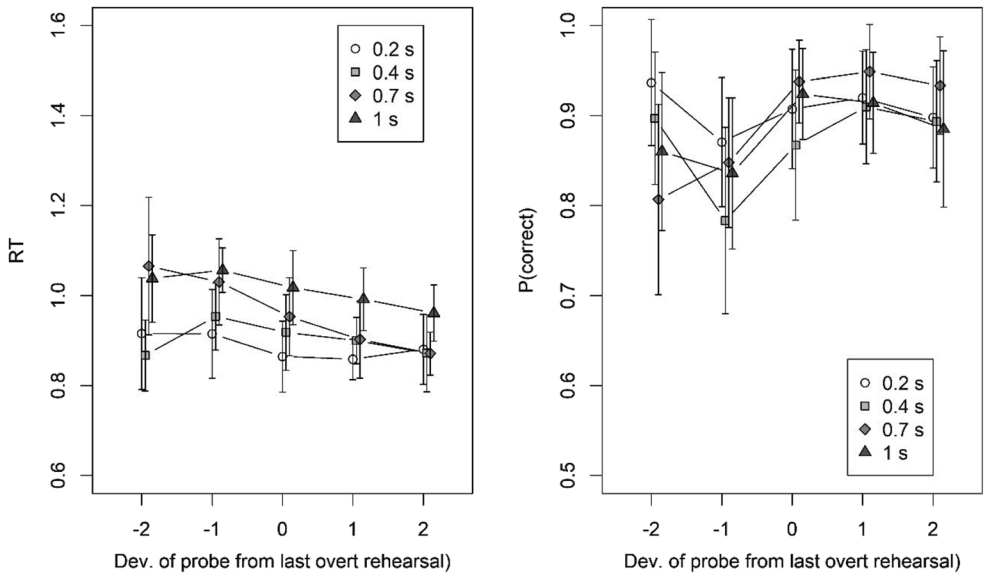

Fig. 13 Mean response times and accuracies of recognition (matching probes) in Experiment 4 as a function of the deviation between the list position that participants have last rehearsed overtly, and the position of the item the probe matched. Positive deviations mean that the probe matched a later list position than the last-rehearsed position. The points legend denotes the pace condition. Error bars are 95% confidence intervals for within-subjects comparisons

attention) could select multiple items simultaneously awaits future investigation.

Although much slower than previously assumed, several pieces of evidence validate the assumption that our adjusted refreshing speed reflects people's intuition regarding the speed at which they could (voluntarily) attend to information in working memory. First, the adjusted speed was similar across several types of memoranda, in line with the assumption that refreshing is domain-general. Second, refreshing speed was faster than articulatory rehearsal speed. Third, articulatory rehearsal was affected by word length, which is a proxy for speech time, whereas refreshing was not. Participants had quite accurate intuitions about the word-length effect on speech time, and their adjusted metronome times tended to closely reflect the overt rehearsal delay imposed by the manipulation of word length. All in all, our experiments point to dissociations in the speed of refreshing and rehearsal in line with the theoretical distinctions usually made between them.

Critics could question the validity of our method of measuring the speed of refreshing by assuming that refreshing is a process outside of people's awareness, or a process that, although aware, proceeds so rapidly that people are unable to monitor it. In that case, people's adjusted metronome speeds would reflect their beliefs about how rapidly they can attend sequentially to items in WM, not the actual speed of refreshing. A recent review distinguished two kinds of refreshing – one slow and deliberate, and the other swift and outside of conscious awareness (Camos et al., 2018). On that assumption, one could argue that our method measures slow, deliberate refreshing but not swift refreshing. There is no logical way to reject this kind of critique because it is impossible to prove the non-existence of an entity. No matter how many attempts to measure swift refreshing fail, proponents of that idea can always argue that the measurement method was not appropriate. At this point, skeptics can only appeal to Occam's razor: We should not assume more entities than we need to explain the data. Whereas there is solid evidence for slow and deliberate refreshing as a process that can be invoked experimentally, and measured as distinct from articulatory rehearsal (Johnson et al., 2005; Raye et al., 2007; Raye, Johnson, Mitchell, Reeder, & Greene, 2002; Souza et al., 2015; Souza et al., 2018), there is no compelling reason to believe that swift refreshing exists.

Our effort to obtain independent behavioral evidence that people actually refreshed in sync with the metronome must be considered a failure. The initially encouraging results we obtained in the studies reported in the Supplementary Online Materials – namely an accuracy benefit for accessing the item that was last to be refreshed according to the metronome – turned out to arise not from people focusing on the items one by one in sync with the metronome. Rather, people only focused on the item indicated by the final state of the metronome when the probe appeared. Once this strategy was prevented in Experiment 3 with the use of a pendulum-like metronome, there was no last-refreshed-item benefit. This negative result stands in contrast to the last-refreshed-item benefit on RTs that Vergauwe and Langerock (2017) have observed. In their procedure, the item to be refreshed at any moment during the retention interval was indicated by highlighting the frame in which that item had been presented. Therefore, participants could have used the same strategy that they apparently used in our Experiments S1 and S2 (Supplementary Online Material): Rather than focusing on each item they were supposed to refresh during the retention interval, they waited until the recognition probe and only then focused on the item in the last highlighted frame. Against that possibility stands the fact that Vergauwe and Langerock observed a last-refreshed-item benefit only on RTs, whereas the beneficial effect we attribute to this strategy was only observed in accuracies. We were not able to reproduce the RT benefit of Vergauwe and Langerock with any of our metronome versions.

We do not know what is responsible for this discrepancy in results – finding out would probably require a long series of experiments comparing the two procedures. We doubt that this would be worth the effort because their difference might reside in subtleties of the procedures, and narrowing them down is irrelevant to answering our question concerning the speed of refreshing.

The results from overt rehearsal give some insights into why our metronome-based method did not work: People found it surprisingly hard to rehearse in sync with the metronome, even at paces for which they expressed high confidence that they could rehearse at them. If the same is true for refreshing, it would explain why people's focus of attention was not systematically at the item they were supposed to refresh at any point in time. The observed speed of rehearsal was somewhat slower than the speed of the metronome. If the same is true for refreshing, the refreshing speed obtained from the adjusted metronome pace would slightly overestimate people's true maximal refreshing speed.

Another possible reason for why we found no evidence for people refreshing in sync with the metronome is that they did not follow our instruction at all – either because they did not understand what we mean by asking them to "think of" the items one by one, or because they had nothing to gain from making the effort. We find this unlikely because all participants did adjust the metronome speed at some point. If they did not at least try to follow the instruction, there would have been no reason for them to do that. Further, they reported high confidence in being able to refresh at paces slower than 200 ms (the common self-adjusted value), and low confidence at paces faster than that (i.e., 100 ms), with intermediate confidence ratings for the pace of 200 ms. This is what we should expect if 200 ms per item is the average maximum speed that people can refresh at: About half the population's maximum

speed is faster than the mean, so they are confident that they can do it; the other half has a slower maximum speed and therefore expresses low confidence. Therefore, the sensitivity of the confidence ratings to the metronome pace validates our estimate of the average maximum speed, and shows that people know what to do when asked to refresh in sync with the metronome. Together with previous studies that have shown that people do understand the instruction to "think of" working memory contents, and benefit from it in working memory tasks (Johnson et al., 2005; Raye et al., 2007; Souza et al., 2015), we believe that the explanation that participants did not understand the instructions, or were not willing to follow them, is very unlikely. Nevertheless, we cannot completely rule out the possibility that participants did not actually refresh the items in sync with the metronome, but rather adjusted the metronome speed according to their subjective theory about their speed of thinking.

In light of these limitations, our conclusion must be tentative and contingent on the assumption that our method measures at least one form of refreshing – one that people can monitor. Refreshing as we measured it provides a faster way to cycle through memory contents than articulatory rehearsal. Nevertheless, it is unlikely that people can sequentially refresh items in working memory at a speed faster than 0.2 s per item. Even this speed might be an overestimation due to people's limited insight into their cognitive processes, as reflected in their overestimation of rehearsal speed. To date, a method for tracking refreshing that does not rely on people's metacognitive abilities remains elusive.

Author note This research was supported by a grant from the Swiss National Science Foundation to K. Oberauer (project 149193). We are grateful to Hannah Schrohe, Isabella Catalano, Julia Hardmeier, Marlies Monch, and Katja Wildmann for their help with collecting the data and coding the overt-rehearsal audio files.

Open Practices Statement All materials (i.e., experimental scripts and stimuli information) and data for all experiments reported here are available at the Open Science Framework at: osf.io/mcxdn; DOI: 10.17605/OSF.IO/MCXDN

References

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavioral Research Methods, Instruments, & Computers, 28*, 584-589.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83-100.

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition* doi:https://doi.org/10.3758/s13421-018-0805-9

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage, 199*, 585-597. doi:https://doi.org/10.1016/j.neuroimage.2019.06.028

Camos, V., Johnson, M., Loaiza, V. M., Portrat, S., Souza, A. S., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences, 1424*(1), 19-32. doi:https://doi.org/10.1111/nyas.13616

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*, 263-276.

Gehring, W. J., Bryck, R. L., Jonides, J., Albin, R. L., & Badre, D. (2003). The mind's eye, looking inward? In search of executive control in internal attention shifting. *Psychophysiology, 40*, 572-585.

Hedge, C., & Leonards, U. (2013). Using eye movements to explore switch costs in working memory. *Journal of Vision, 13*, 1-19. doi: https://doi.org/10.1167/13.4.18

Hedge, C., Oberauer, K., & Leonards, U. (2015). Selection in spatial working memory is independent of perceptual selective attention, but they interact in a shared spatial priority map. *Attention, Perception & Psychophysics, 77*, 26653-22668. doi:https://doi.org/10.3758/s13414-015-0976-4

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect. *Memory & Cognition, 39*, 1198-1210. doi:https://doi.org/10.3758/s13421-011-0094-z

Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience, 4*, 268-280.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: Prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective, and Behavioral Neuroscience, 5*, 339-361.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual an Motor Skills, 15*, 646.

Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2018). What is the time course of working memory attentional refreshing? *Psychonomic Bulletin & Review, 25*, 370-385. doi:https://doi.org/10.3758/s13423-017-1282-z

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), The Psychology of Learning and Motivation Vol. 46pp. 155-200. San Diego: Academic Press.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor (Version 0.9.12.2). Retrieved from http://cran.at.r-project.org/web/packages/BayesFactor/index.html

Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50*(4), 257-269.

Oberauer, K. (2019). Working memory and attention - a conceptual analysis and review. *Journal of Cognition, 2*, 1-23. doi:https://doi.org/10.5334/joc.58

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review, 18*(1), 10-45.

Oberauer, K., Wendland, M., & Kliegl, R. (2003). Age differences in working memory - The roles of storage and selective access. *Memory & Cognition, 31*(4), 563-569.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8-13. doi:https://doi.org/10.1016/j.jneumeth.2006.11.017

Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. A., & Greene, E. J. (2002). Neuroimaging a single thought: Dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage, 15*, 447-453.

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43*, 135-145.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 419-425.

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences, 1339*, 20-31. doi:https://doi.org/10.1111/nyas.12603

Souza, A. S., Rerko, L., & Oberauer, K. (2016). Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 890-910. doi:https://doi.org/10.1037/xhp0000192

Souza, A. S., Vergauwe, E., & Oberauer, K. (2018). Where to attend next: guiding refreshing of visual, spatial, and verbal representations in working memory. *Annals of the New York Academy of Science* doi:https://doi.org/10.1111/nyas.13621

Tanoue, R. T., & Berryhill, M. E. (2012). The mental wormhole: Internal attention shifts without regard for distance. *Attention, Perception, & Psychophysics* doi:https://doi.org/10.3758/s13414-012-0305-0

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). Revisiting the attentional demands of rehearsal in working-memory tasks. *Journal of Memory and Language, 105*, 1-18. doi:https://doi.org/10.1016/j.jml.2018.10.005

Vergauwe, E., & Cowan, N. (2014). A common short-term memory retrieval rate may describe many cognitive procedures. *frontiers in human neuroscience, 8*. Retrieved from doi:https://doi.org/10.3389/fnhum.2014.00126

Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory and Language, 96*, 23-35. doi:https://doi.org/10.1016/j.jml.2017.05.001

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1072-1095.

Vergauwe, E., Hardman, K. O., Rouder, J. N., Roemer, E., McAllaster, S., & Cowan, N. (2016). Searching for serial refreshing in working memory: Using response times to track the content of the focus of attention over time. *Psychonomic Bulletin & Review* doi:https://doi.org/10.3758/s13423-016-1038-1

Vergauwe, E., Langerock, N., & Cowan, N. (2018). Evidence for spontaneous serial refreshing in verbal working memory? *Psychonomic Bulletin & Review, 2*, 674-680. doi:https://doi.org/10.3758/s13423-017-1387-4

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. L. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavioral Research Methods, 41*, 534-538. doi:https://doi.org/10.3758/BRM.41.2.534

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.