Online Supplementary Materials

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. Original research manuscript. *Annals of the New York Academy of Sciences.*

This file contains detailed methods descriptions and supplementary analyses from:

Experiment 1

* Participants, Materials, and Procedure
* Mixture Model Results

Experiment 2

* Participants, Materials, and Procedure
* Mixture Model Results

Experiment 3

* Participants, Materials, and Procedure
* Mixture Model Results

References

Figures

* Figures S1-S4

# Experiment 1

## Method

### Participants

Thirty-nine students from the University of Zurich (28 women; mean age = 24.8 years) took part in Experiment 1 (Exp. 1A *n* = 18; Exp. 1B, *n* = 21). One participant from Experiment 1B was excluded due to a misunderstanding of the refreshing instruction. For all experiments reported in this paper, participants were compensated with course credit or 15 Swiss francs per hour, they read and signed an informed consent form prior to the study, and were debriefed at the end. The research protocol was approved by the Ethics Committee of the Philosophical Faculty of the University of Zurich.

### Materials and Procedure

All experiments were programmed using MATLAB and the Psychophysics Toolbox extension 1,2.

Participants performed a color reconstruction task 3,4. At the beginning of each trial, a white fixation cross was shown against a grey background for 500 ms, followed by the presentation of six colored disks for 1000 ms. The colors of the memoranda were sampled from a color wheel with 360 values evenly distributed on the hue dimension in the HSL (*hue, saturation, and lightness*) color model, with saturation = 1 and lightness = 0.5. Colors were selected randomly with the constrain that all six colors were at a minimum distance of 20° of each other on the color wheel. The colored disks had to be maintained over a retention interval (RI) of 2.5 s, after which a test display was shown containing a color wheel (randomly rotated from trial to trial), a white circle frame (indicating the location of the target item), and a question (“*Color?*” in German). Participants were asked to indicate the color of the target by clicking on a point on the color wheel. The next trial started 1 s later. Instructions emphasized accuracy but not speed. Participants were asked to repeat the sequence *"der-die-das"* aloud throughout the experiment to minimize the chance of verbal encoding and articulatory rehearsal.

The color wheel provides a fine-grained, approximately continuous response scale that yields a quantitative estimate of retrieval error, yielding a sensitive measure of memory accuracy. A further advantage is that the error distributions can be submitted to a mixture model to obtain a more detailed picture of experimental effects (see below). The position of the color wheel was varied randomly on each trial to ensure that any type of response tendencies (e.g., always guess by clicking on a specific color-wheel location; or to always click on the color close in space to the location of the target) yields a random color selection and therefore can be more easily distinguished from true recall attempts. Moreover, this helps ruling out possible contributions of spatial memory to the retention of the colors.

We inserted a refreshing manipulation during the RI: Four central arrow cues were presented sequentially, each for .5 s. We will refer to the presentation of each (retro-)cue as a refreshing step. Participants were instructed to think of the item the arrow pointed to. Participants were also instructed that the cues did not reliably indicate the item to be tested, but they were urged that thinking of the cued item is part of their main task (see Figure 1 in the main article).

There were five possible sequences of cues. Assume that the letters ABCD represent four diferrent randomly selected memory items: The four arrows could point to four different items (A-B-C-D), they could point to two different items once and to a third item twice (A-B-A-C; A-B-C-B; A-B-C-A), or to two different items twice (A-B-A-B). Note that across two sucessive cues, different items were always cued. Given these constrains, there were three items that were not cued (0-Refreshing items), two items that were cued once (1-Refreshing items), and one item that was cued twice (2-Refreshing items), on average, in each memory array. The cue sequence for each trial was chosen at random.

Items refreshed twice were necessarily refreshed once in cue position 1 or 2, and then again in cue position 3 or 4. In contrast, 1-Refreshing items could have been refreshed only once in any position. Therefore, compared with 1-Refreshing items, 2-Refreshing items have either a shorter interval between memory-display offset and the first refreshing, or between last refreshing and test-display onset. Across Experiments 1A (n = 18) and 1B (n = 20), we matched the 1-Refreshing and 2-Refreshing targets regarding one of these intervals.

In each experiment, participants performed 390 trials which were equaly and randomly split into the three refreshing levels. In the beginning of the session, participants completed 3 practice trials that were discarded from subsequent analyses.

## Mixture Model Results

Our main dependent variable was recall deviation. We ran a Bayesian ANOVA (BANOVA) using the absolute recall deviation in each trial as the predicted variable and refreshing level and experiment as the predictor variable. This analysis indicated that items refreshed more often were better recalled (see BF values in Table 1 in the main text).

We have also fitted to the distribution of recall deviations a probabilistic mixture model 4,5. The mixture model estimates the probability of recalling the target, and the variability (i.e., precision) with which the target's color is recalled, given that it is recalled. In case the target was not recalled, this model assumes that observers either guess by randomly selecting any color in the color wheel, or that they erroneously retrieved one of the non-target items (transposition error). We also fitted a more constrained mixture model that does not include transposition errors 4, but this model yielded a lower fit to the data than the non-constrained mixture model, as indexed by the Bayesian Information Criterion (BIC).

Here, we present the effect of refreshing on the probability of recalling the target, computed as 1 – guessing rate – transposition rate, and on the precision parameter, which is the standard deviation in the mixture model (see Figure S1).

As shown in Figure S1a, the probability of recalling the target increased with refreshing frequency. The BANOVA indicated that the best model included both the main effects of refreshing level (0, 1, and 2) and Experiment, and no interaction (BF over the Null = 6.75). The regression with refreshing frequency as a linear predictor yielded a BFLinear/Null = 775, indicating a strong linear trend. Figure S1b shows the precision of recalling the target color. There were no systematic differences between refreshing conditions on this measure. The BANOVA results indicate that the Null model should be preferred over all alternative models by a factor of at least 5 (linear regression BFLinear/Null = 2 x 10-4).

We also analyzed the probability of guessing and the probability of transposition errors separately, but in isolation, none of these parameters showed the linear improvement in performance as a function of refreshing that was observed in the raw data, and in the probability of recall.

In sum, the refreshing benefit observed in the raw data was reflected in an increase in the probability of recalling the target item estimated from the mixture model, but not in the precision of recall.

# Experiment 2

## Method

### Participants, Materials, and Procedure

A new sample of 24 participants (18 women; average age: 24.4 years) took part in Experiment 2. Participants completed the Refreshing condition in one session and the two Baseline conditions in another session. Order of sessions was counterbalanced across participants.

Figure S2 shows the events and timing in each condition of Experiment 2. The *Refreshing condition* was similar to the one in Experiment 1, with three exceptions. First, 1-Refreshing targets were selected from any cue position in the cue sequence. Second, the pre-cue interval was increased from .5 s to 1 s; and third, the inter-trial interval (ITI) was increased to 7.5 s. The baseline session was divided into two conditions: the *Baseline-long* and the *Baseline-short* condition. As shown in Figure S2, in the Baseline-long, RI was 3 s, and ITI was 7.5 s. Thus, the Baseline-long condition is matched to the Refreshing condition with regard to overall RI and ITI. In the Baseline-short condition, RI was 1 s, which was equal to the interval between memory array offset and the first refreshing cue in the refreshing condition. For this condition, ITI was 1 s. We varied RI and ITI in sync to hold temporal distinctiveness of successive trials constant across all conditions. Temporal distinctiveness is defined as the ratio of two time intervals: (1) the time between encoding of the current memory array and the current test, and (2) the time between encoding of the array in the preceding trial and the current test 6; see Figure S2 for the time ratios).

The Refreshing and Baseline sessions comprised 240 trials each. Each refreshing level had 80 trials (randomly intermixed). Each baseline condition comprised 120 trials, which were blocked, and the order of blocks was counterbalanced across participants. Before each block, participants completed 6 practice trials to get used to the task requirements (refreshing manipulation or not) and the duration of the RI and ITI in that block.

## Mixture Model Results

We fitted the probabilistic mixture model to the data of this experiment (see Figure S3). The probability of recalling the target increased as the frequency of refreshing steps directed at the target increased (Figure S3a). The BANOVA yielded a BFRef/Null = 7.9, and the linear regression yielded a BFLinear/Null = 2.1 x 105. Comparison of the two baseline conditions did not favor the alternative hypothesis (BFBase/Null = 1.82). Finally, the comparison of the Baseline-short to the 0-Refreshing target condition yielded a BF = 0.32; indicating that the Null should be favored by a factor of about 3.

There were no consistent differences between refreshing conditions in recall precision (BFRef/Null = 0.13; see Figure S3b). For precision the BF also favored the Null for the comparison of the two baseline conditions with each other (BFBase/Null = 0.79). The comparison of the Baseline-short to the 0-Refreshing targets also favored the null (BF = 0.41).

# Experiment 3

The goal of Experiment 3 was to examine whether totally non-predictive cues can be used to guide the pattern of refreshing.

## Method

### Participants, Materials, and Procedure

A new sample of 28 participants (18 women; average age: 24.3 years; range 19-35 years) took part in Experiment 3. Participants completed two sessions comprising trials with the color reconstruction task and the refreshing manipulation used in Experiment 2. The task was similar to the one in Experiment 2 with two exceptions. First, the color wheel used the CIE L\*a\*b color space as described by Zhang and Luck (2008). Second, the target of recall was randomly selected in each trial, and all items in the memory array were equally likely to be tested. Hence, in average, the probability of testing a 0-, 1-, or 2-Refreshing item was of 3/6, 2/6, and 1/6, respectively, thereby matching the number of items that were targets of 0-, 1-, and 2-Refreshing steps, on average, in a trial. This contrasts with the probability of testing 0-, 1-, and 2-Refreshing items in Experiments 1 and 2, which was always 1/3.

Participants completed 600 trials, which were evenly split into two 1-hour sessions. We doubled the number of experimental trials to ensure that even with the reduced chance of testing 1- and 2-Refreshing targets, there would be a sufficiently large number of trials in each refreshing condition. Five participants completed only one session and therefore contributed with only half of the data to the analysis.

## Mixture Model Results

We also fitted the mixture model to the data of Experiment 3 (see Figure S4). Visual inspection of this figure suggest that the effect of refreshing was reflected in an increased probability of recalling the target but not on the precision parameter. For probability of recall, a model testing for a main effect of refreshing level yielded a BFRef/Null = 1.57, yielding only weak evidence for an effect of refreshing on probability of recall. The linear regression captured evidence for an effect of refreshing, BFLinear/Null = 1.4 × 108. For precision, both the model with a main effect of refreshing frequency (BFRef/Null = 0.38) and the linear regression (BFLinear/Null = 0.18) favored the Null model.

# References

1. Brainard, D. H. 1997. The psychophysics toolbox. *Spat. Vis.* **10:** 433–436

2. Pelli, D. G. 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10:** 437–442

3. Wilken, P. & Ma, W. J. 2004. A detection theory account of change detection. *J. Vis.* **4:** 11

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

5. Bays, P. M., Catalao, R. F. G. & Husain, M. 2009. The precision of visual working memory is set by allocation of a shared resource. *J. Vis.* **9:** 7

6. Brown, G. D. A., Neath, I. & Chater, N. 2007. A temporal ratio model of memory. *Psychol. Rev.* **114:** 539–576

# Figure Captions

*Figure S1.* Panel a shows the probability of recalling the target item, and panel b shows the precision with which the target was recalled, estimated from the mixture model fitted to the data of Experiments 1A and 1B. Error bars represent 95% within-subjects confidence intervals.

*Figure S2.* Conditions in Experiment 2. M = memory display; T = test display; R = refreshing step; event durations are shown above each event marker. For the test display the duration is the predicted mean reaction time to recall the target color. The memory and test displays in trial n and n-1 are depicted in black and grey, respectively. The panel also shows the estimated temporal distinctiveness of trials.

*Figure S3.* Results of the mixture model fitted to the data of Experiment 2.

*Figure S4.* Results of the mixture model fitted to the data of Experiment 3.

C:\Users\Alessandra\Dropbox\Writing papers\Rehearsal project\Refreshing Experiments\plots\submission plots\Souza_FigureS2.tiff

*Figure S1.* Panel a shows the probability of recalling the target item, and panel b shows the precision with which the target was recalled, estimated from the mixture model fitted to the data of Experiments 1A and 1B. Error bars represent 95% within-subjects confidence intervals.

C:\Users\Alessandra Souza\Dropbox\Writing papers\Rehearsal project\Refreshing Experiments\paper ThinkColor\newSubmission\submission plots\Souza_FigureS2.tif

*Figure S2.* Conditions in Experiment 2. M = memory display; T = test display; R = refreshing step; event durations are shown above each event marker. For the test display the duration is the predicted mean reaction time to recall the target color. The memory and test displays in trial n and n-1 are depicted in black and grey, respectively. The panel also shows the estimated temporal distinctiveness of trials.

*C:\Users\Alessandra Souza\Dropbox\Writing papers\Rehearsal project\Refreshing Experiments\paper ThinkColor\newSubmission\submission plots\Souza_FigureS3.tiff*

*Figure S3.* Results of the mixture model fitted to the data of Experiment 2.

C:\Users\Alessandra Souza\Dropbox\Writing papers\Rehearsal project\Refreshing Experiments\paper ThinkColor\newSubmission\submission plots\Souza_FigureS4.tiff

*Figure S4.* Results of the mixture model fitted to the data of Experiment 3.