Research Article

No evidence that self-rated negative emotion boosts visual working memory precision

Alessandra S. Souza

University of Zurich, Switzerland, University of Porto, Portugal

Theresa Thaler

University of Zurich, Switzerland

Heinrich R. Liesefeld

Ludwig-Maximilians-Universität München, Germany

Flávia H. Santos

University College Dublin, Ireland

Débora S. Peixoto & Pedro B. Albuquerque

University of Minho, Portugal

Accepted for publication in

Journal of Experimental Psychology: Human Perception and Performance

©American Psychological Association, [2020]. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: <http://dx.doi.org/10.1037/xhp0000891>

Word-count: 19'440

Running-head: Emotion & Working Memory

Author Note

Alessandra S. Souza, Department of Psychology, University of Porto, Portugal. Theresa Thaler, Department of Psychology, University of Zurich, Switzerland. Heinrich R. Liesefeld, Department Psychologie, Ludwig-Maximilians-Universität München, Germany. Flavia H. Santos, School of Psychology, University College Dublin, Ireland. Débora S. Peixoto and Pedro B. Albuquerque, Department of Basic Psychology, University of Minho, Portugal. Research reported here was supported by a grant from the Swiss National Science Foundation (grant 100019_169302) to A. S. Souza, a grant from the German Research Foundation (DFG; grant LI 2868/3-1) to H. R. Liesefeld. F. H. Santos was funded by Portuguese National Funding Agency for Science, Research and Technology (FCT-SFRH/BPD/91347/2012). Experiment 5 was conducted at the Psychology Research Centre (PSI/01662), School of Psychology, University of Minho and supported by the Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education, through the State Budget (UID/PSI/01662/2019).

Correspondence should be addressed to Alessandra S. Souza, Faculty of Psychology and Education Sciences, University of Porto, Rua Alfredo Allen, 4200-392 Porto, Portugal. E-mail: alessandra@fpce.up.pt

OSF link: <https://osf.io/wzvmtx>

Abstract

Emotion is assumed to change how people process information by modulating attentional focus. Two recent studies (Spachholz, Kuhbandner, & Pekrun, 2014; Xie & Zhang, 2016) reported that self-reported negative emotion boosted the precision with which information was stored in visual working memory. Here we attempted and failed to replicate these findings across 7 studies conducted in four countries. Emotion was induced by presenting emotional images (negative, neutral, and positive) before each trial of a visual working memory task (6 experiments) or the images were combined with emotional music during a 3-min induction phase (1 experiment) occurring prior to the memory task. In the visual working memory task, participants stored (emotionally-neutral) continuously-varying colored dots or oriented triangles. At test, the color or orientation of a probed item was reproduced. Although participants reported changes in their emotional state commensurate with the manipulations, six experiments showed substantial evidence against changes in visual working memory precision (and quantity) under negative (and positive) emotion in comparison to neutral; whereas one condition, in one study, showed increased precision under both negative and positive emotion compared to neutral. These results challenge the view that emotion modulates visual working memory quality and quantity.

Keywords: attention, emotion, memory fidelity, visual working memory, continuous memory

Public Significance Statement

In our everyday life, we are constantly exposed to stimuli that elicit brief emotional reactions, for example, sad news, exciting images, or melancholic music. This study examined the possibility that changes in emotional state produced by these emotional stimuli affects how other (neutral) visual information is stored in mind over brief intervals. Our results provide substantial evidence that emotions do not bias visual memory over the short-term, suggesting that the fidelity and quantity of our memory for neutral information is not affected by the moment-to-moment fluctuations in our emotions.

No evidence that self-rated negative emotion boosts visual working memory precision

In a variety of daily tasks, we need to keep in mind visual information that is no longer in view, but is needed for completing our goals. For example, to play ball games – such as volleyball, basketball, or soccer – we need to keep track of the ball and of the position of the other players. If we fail to keep this information in mind, we cannot block the adversary passes or pass the ball to a team-mate. Visual working memory is the memory system that maintains active the visual information we need for ongoing activities. The capacity of visual working memory is severely limited: people can only hold in mind a small number of visual objects with imperfect fidelity (Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 2013). This capacity limit provides an upper bound on our ability to effectively act in the visual world around us. Accordingly, researchers have been interested in understanding the factors that influence the quantity and quality of the visual information stored in this system.

Recently, the impact of normal fluctuations of emotional states upon visual working memory has gained renewed interest. In our everyday life, we are regularly confronted with situations or circumstances that cause an emotional reaction (Kensinger, 2007). Those emotional responses are assumed to change the way we value our experiences, and how we perceive, remember, and react to the environment around us (Dolan, 2002; Gray, 2004; Okon-Singer, Hendler, Pessoa, & Shackman, 2015). Given that working memory is required to deal with incoming information, to integrate it with the broad knowledge-basis of an individual, and

for the selection of actions (Oberauer, 2009), it is not surprising that there is interest in understanding how emotion may change processing priorities within this system.

Emotion and Visual Working Memory

There are two routes to study the impact of emotion on cognition (Ribeiro, Santos, & Albuquerque, 2019). One possibility is to assess how emotionally-charged information is processed in comparison to non-emotional (neutral) information. This implies that the presence of emotional stimuli might bias people's perceptions and memories.

Studies that presented emotional visual stimuli as the memoranda in visual working memory tasks observed either improvements or impairments in performance compared to the processing of neutral information. For example, in the study of Jackson, Linden, and Raymond (2014), participants had to memorize faces for a recognition test. Angry faces improved recognition performance compared to neutral faces (see also Jackson, Wolf, Johnston, Raymond, & Linden, 2008; Jackson, Wu, Linden, & Raymond, 2009; Thomas, Jackson, & Raymond, 2014). In the study of Kensinger and Corkin (2003), conversely, processing of fearful faces in a 2-back working memory task was associated with slower responses compared to neutral faces, showing a detrimental effect of emotional processing for task performance (see also Bergmann, Rijpkema, Fernández, & Kessels, 2012; Brose, Schmiedek, Lövdén, & Lindenberger, 2012; Curby, Smith, Moerel, & Dyson, 2019). The effect of emotion in these conditions, however, could be due to the emotional stimuli having low-level properties that are more salient than the ones associated with the neutral stimuli (Constant & Liesefeld, 2020; Jingling & Zhaoping, 2008). This would indicate that emotion has an influence via changes in

bottom-up processes that attract attention to these stimuli leading to their preferential processing at expense of other (neutral) stimuli. This is in line with several models of emotion-cognition interactions proposing that emotional stimuli have a competitive advantage in grabbing cognitive resources for their processing, such as the *dual competition framework* (Pessoa, 2009) or the *arousal-biased competition* model (Mather & Sutherland, 2011).

A second way to assess the impact of emotion on cognition is to examine how variations in emotional state change the way people react to the same set of neutral information. For example, when stressed or sad, people may have a different processing style and use their cognitive capacities differently than when they are under an elated or relaxed mood. In this scenario, emotion would change the engagement of cognitive processes and/or brain regions, yielding benefits or costs to observed performance.

In line with this possibility, three recent studies indicated that negative affective state (induced by experimental manipulations and validated by self-report) increased the priority for the storage of details in visual working memory. In the study of Spachtholz et al. (2014), two groups of participants were exposed to an autobiographic task inducing either a neutral or a negative affective state. Afterwards, they completed a visual working memory task in which a set of colored dots were memorized. At test, the location of one dot was highlighted, and participants reproduced its color using a continuous color wheel. Responses in this task can be modeled to extract parameters related to the probability of correct retrieval of the tested item from working memory (as opposed to guessing) and the imprecision with which this information was stored (Zhang & Luck, 2008). By modeling their data, Spachtholz et al. (2014) observed that, in comparison to the neutral-state group, participants in the negative condition

showed a lower probability of retrieving the correct color, but when this color was retrieved, it was reproduced with higher precision. These results suggest that a self-rated negative emotional state changes the trade-off between the quantity and the quality of visual working memory representations.

In the study of Xie and Zhang (2016), affective states were manipulated on a trial-by-trial basis by presenting emotional images drawn from the International Affective Picture System, IAPS (Lang, Bradley, & Cuthbert, 2008). Images were selected that were associated with positive, neutral, and negative states. In each trial, an image was presented for 4 s while participants assessed the emotion it elicited on them using a 9-point valence scale. Thereafter, a set of colored dots (or closed-contour shapes) was presented for memorization, and after a brief interval, one of them was cued to be reproduced using a color (or shape) wheel. Again, modeling of the data in this task indicated that responses in trials in which negative images were presented were more precise, but this time with no change in the probability of retrieval of the tested item, compared to trials with neutral and positive images (which did not significantly differ). Unlike the study by Spachtholz et al. (2014), these results indicate only a boosting effect of self-rated negative emotion on the fidelity of visual working memory representations with no cost to the quantity of the information stored.

Finally, in a recent study by Long et al. (2020), a similar induction procedure as the one used by Xie and Zhang (2016) was used: an IAPS image was presented for rating before each visual working memory trial. The memoranda consisted of four oriented gratings (aka Gabor stimuli). They observed that when the presentation duration of the memoranda was short (200 ms), inducing either a negative or neutral state had no impact on memory performance,

whereas when the presentation duration was longer (500 ms), negative emotion induced a trade-off: memory precision was higher, but the probability of recall was lower – similarly to the findings of Spachtholz et al. (2014). They interpreted their findings as suggestive that negative emotion affects late stages of the consolidation processes in visual working memory, that need time to be deployed. The moderation of the negative-emotion effect by presentation duration in Long et al., however, cannot account for the variability in outcomes between the study of Spachtholz et al. (2014) and Xie and Zhang (2016). Spachtholz et al. used a 200-ms presentation duration and reported a tradeoff between quantity and quality, whereas for this duration Long et al. (2020) found no effect of emotion. Xie and Zhang (2016), in contrast, used presentation durations of 400 ms (Exp. 1) and 1000 ms (Exp. 2), and only observed a precision boost.

In sum, these three studies pointed to an improvement in memory quality under self-rated negative emotion. While in Spachtholz et al. (2014) and Long et al. (2020), this fidelity-boost was accompanied by a decrease in the number of items stored, there were no costs to the number of items stored in Xie and Zhang (2016). What explains this discrepancy is not clear. Xie and Zhang (2016) argued that the results of Spachtholz et al. (2014) might reflect a bias (participants choosing to favor more precise representations instead of storage of all memory items), whereas their own study pointed to a true emotional boost. Regardless of the reasons for their discrepant results regarding memory quantity, these studies converged on indicating higher memory precision under self-rated negative emotion. Xie and Zhang (2016) argued that this boost might be due to negative emotion recruiting brain regions associated with emotional processing that would otherwise not be recruited by the task, such as the amygdala-

hippocampus circuit. The modulations observed were assumed to be in line with the idea that negative emotion changes the state of the cognitive system towards more detail-oriented processing (Schwarz & Clore, 2003).

One major advantage of manipulating emotion independently of the stimuli used in the task is that it removes confounds of low-level feature differences between emotional and non-emotional stimuli as an explanation for performance differences between conditions. Accordingly, this type of manipulation provides a direct window on emotion-cognition interactions that goes beyond differential competition for processing afforded by the type of stimuli encountered. Here, we were interested in assessing how emotional states could modulate processing of otherwise neutral information in visual working memory, following the studies of Xie and Zhang (2016), Long et al. (2020), and Spachtholz et al. (2014).

The Present Study

The results of Xie and Zhang (2016), Long et al. (2020), and Spachtholz et al. (2014) suggest that normal fluctuations in emotional states – similar to the ones commonly experienced by people during their daily lives – can modulate processing priorities within visual working memory. The question that interested us when we started this project was to understand the reasons for this emotional boost. One hypothesis that we considered was that emotion could modulate the engagement or efficiency of attentional processes within working memory. For example, negative emotion could up-tune attentional resources, thereby allowing participants to encode the information better, maintain it in mind, and/or select it from within visual working memory.

Emotional states have been long assumed to modulate attentional processing, with negative affect assumed to narrow attentional focus and positive affect assumed to broaden it (Easterbrook, 1959; Fredrickson & Branigan, 2005; Gable & Harmon-Jones, 2008; Uddenberg & Shim, 2015). Accordingly, we reasoned that if negative emotion served as an alerting signal increasing the general propensity of the system to process information (Haupt, Ruiz-Rizzo, Sorg, & Finke, 2019), it may also increase attentional effects observed in visual working memory tasks or it may allow for better protection against time-based forgetting or interference at recall (Tabi, Husain, & Manohar, 2019). With this reasoning in mind, we ran Experiment 1 in Switzerland, Experiment 4 in Germany, and Experiment 5 in Portugal. The goal of these studies was originally to replicate the findings of Xie and Zhang (2016) and Spachholz et al. (2014) and to examine its potential interaction with either an attentional manipulation (Experiments 1 and 5) or with a manipulation of retention interval (Experiment 4). Across these three studies, however, we were unable to replicate the finding that self-rated negative emotion as opposed to neutral and positive emotion boosted the storage of details in visual working memory. We then ran additional studies with the sole goal of trying to replicate the finding that negative emotion increased memory precision (Experiments 2a, 2b, and 3). Given these failed replication attempts, we ran a last preregistered (<https://osf.io/7jhgx>) online study with a large sample of participants ($N = 151$) contrasting conditions purportedly inducing negative and neutral emotional states and addressing potential issues with our previous replication attempts that were highlighted during the review process.

Table 1 presents an overview of the studies included in the paper and of the three studies available in the literature. Six out of the seven experiments reported in the present

paper showed substantial evidence against an effect of self-rated negative emotion on any parameter (either quality or quantity) of visual working memory performance, and only one condition in one study (Experiment 4) showed higher memory precision for both negative and positive emotion compared to neutral.

In conjunction, our results show that the effects of negative emotion on the processing of visual information in working memory are not as robust as the initial reports may have suggested. We discuss several reasons why we failed to replicate these results, and urge for more systematic investigations on the potential impact of fluctuations in emotional state upon visual working memory.

All data, analysis scripts, and supplementary materials are available at the Open Science Framework: <https://osf.io/wzvmx>

Experiment 1

This experiment was modeled after the study of Xie and Zhang (2016, Experiment 1) in which the emotional state of the participants was manipulated via the presentation of emotional images selected from the IAPS on a trial-by-trial basis. Participants were presented with either a negative, neutral, or positive image for rating prior to the presentation of each trial of a visual working memory task involving the encoding of five colored dots, and the reproduction of one of these colors using a continuous color wheel.

Table 1

Overview of the Features of the Experiments conducted here (E1-E6) and the ones reported in the Literature (SP, XZ, and LO).

Exp.	N	Country	Emotion Induction	Induc. Dur	Rating Scale	Memory Stimuli	Memory Load	Mem. Dur	Trials	Additional manipulation	Valence Effect on Precision
SP	80	DE	Autobiogr. Mem. + Emotional Music Between-subjects	3 min	9	Colored squares	4	0.2 s	96	Iconic memory test	Neg>Neu
XZ1	18	USA	Images IAPS, blocked	4 s	9	Colored squares	5	0.4 s	120	---	Neg>[Neu&Pos]
XZ2	19	USA	Images IAPS, trial-by-trial	4 s	9	Continuous shapes	4	1 s	120	---	Neg>[Neu&Pos]
LO1	30	CN	Images IAPS, blocked	4 s	9	Oriented gratings	4	0.2 s	120	---	No
LO2	30	CN	Images IAPS, blocked	4 s	9	Oriented gratings	4	0.5 s	120	---	Neg > Neu
LO3	30	CN	Images IAPS, blocked	4 s	9	Oriented gratings	4	varied	120	Mem. Dur = 0.2 or 0.5 s	Neg > Neu ¹
E1	22	CH	Images IAPS, trial-by-trial	Max. 6 s	9	Colored dots	5	0.4 s	120	Retro-cue	No
E2a	20	CH	Images IAPS, trial-by-trial	Max. 6 s	9	Colored dots	5	0.4 s	120	---	No
E2b	20	CH	Images IAPS, trial-by-trial	4 s	9	Colored dots	5	0.4 s	120	---	No
E3	42	CH	Images IAPS, trial-by-trial	4 s	9	Oriented triangles	4	0.4 s	120	---	No
E4	17	DE	Images IAPS, trial-by-trial	3 s	5	Colored squares	5	0.4 s	100	Retention (1 vs 3 s)	[Neg&Pos]>Neu ²
E5	60	PT	Images IAPS + Emotional Music between-subjects	3 min		Colored dots	6	1 s	75	Retro-cue	No
E6	151	Online USA	Images IAPS, trial-by-trial	4 s	9 points	Oriented triangles	4	0.4 s	120	---	No

Note. XZ1 and XZ2 = Experiments 1 and 2 of Xie and Zhang, 2016; SP = Spachtholz et al., 2014; LO1 to LO3 = Experiments 1 to 3 of Long et al., 2020. ¹ Effect present only in the 0.5 s condition. ² Effect present only in the 1-s retention interval condition. Trials refer to the number of trials per design cell.

In addition to the manipulation of emotion, in half of the trials, we included an attention manipulation during the retention interval of the visual working memory task. The attentional manipulation consisted of the presentation of an arrow which pointed to the location of one of the memory items. The arrow indicated which memory item would be tested at the end of the trial, and participants were urged to use the cue for preparing for the test. The arrow served therefore as a so-called “*retro-cue*” (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003). Retro-cues guide attention to the information that is relevant for the upcoming memory test, and they reliably yield better performance in visual working memory tasks (for a review see Souza & Oberauer, 2016). Retro-cues have been shown to improve working memory recall for several reasons: (a) the memory representation relevant to the test can be selected among the other working memory contents without interference yielded by the memory test (Makovski, Sussman, & Jiang, 2008; Souza, Rerko, & Oberauer, 2016), (b) it allows the irrelevant (non-cued) information to be removed from working memory thereby freeing capacity (Souza, Rerko, & Oberauer, 2014; Williams, Hong, Kang, Carlisle, & Woodman, 2013), and (c) it strengthens the representation of the attended (cued) item (Souza & Oberauer, 2017; Souza, Rerko, & Oberauer, 2015).

Our original intention was to assess whether self-rated negative emotion (and positive emotion) altered the way in which the retro-cues were used to improve performance in the visual working memory task.

Method

Participants

Xie and Zhang (2016) reported a $d_z = 0.58$ with a sample-size of $N = 18$. A power analysis performed in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) for a one-tailed t -test indicated that a sample size of $N = 20$ yields a power of $1 - \beta = .80$ ($\alpha = .05$) to detect an effect of $d_z = 0.58$.

Given that we used Bayesian inference in our analyses (Schönbrodt & Wagenmakers, 2018), our decision for Experiments 1-3 was to start with a sample-size close to the one used in Xie and Zhang (2016) and, if needed, increase it for obtaining substantial evidence - i.e., a Bayes Factor of $BF > 3$ (Jeffreys, 1961; Kass & Raftery, 1995; Wetzels & Wagenmakers, 2012) - for or against our hypotheses. Increasing the sample size after seeing the data is not problematic in Bayesian inference because this analysis gathers evidence for and against a hypothesis; and therefore this practice does not unduly inflates false positives (Rouder, 2014; Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017).

Twenty-two students (16 women; mean age = 23.5 years) of the University of Zurich took part in two 1-hour sessions in Experiment 1. Participation was compensated with either 30 Swiss Francs (ca. 30 US dollars) or course credit. Students reported normal color vision and visual acuity.

For all experiments reported in this paper, participants signed an informed consent form prior to study and received a debriefing after the data collection was finished. The Ethics Committee of the Faculty of Arts of the University of Zurich approved the study protocol

(approval number 18.6.7) used for all experiments conducted in/from Switzerland (Exps. 1-3 and 6).

Participants in Experiments 1-2 completed a State-Trait-Anxiety-Depression Inventory (STADI; Laux, Hock, Bergner-Köther, Hodapp, & Renner, 2013) before starting the experiment to ensure that they did not suffer from depression or anxiety. Descriptives of the inventory responses are provided on the OSF (<https://osf.io/wzvmx/>).

Stimuli

The emotional states were induced with pictures from the IAPS (Lang et al., 2008). For each valence condition (i.e., negative, neutral, and positive), 120 images were selected. The images were chosen based on the images used both by Xie and Zhang (2016) – which only used 60 images per valence condition – and Bermudez and Souza (2017). We had to include additional images because our experiment comprised twice as many trials overall (due to our added manipulation of cueing) compared to the study of Xie and Zhang (2016).

The cutoff to classify images as belonging to the different valence conditions was based on the normative ratings obtained on a 9-point valence scale that ranges from very negative to very positive (Lang et al., 2008). The selected negative images had ratings between 1 and 4.5 ($M = 2.76$), neutral images had ratings between 4.5 and 5.5 ($M = 5.0$), and positive images had ratings between 5.5 and 9 ($M = 7.24$). For arousal, negative ($M = 5.14$) and positive images ($M = 5.07$) were matched to have similar levels, with both having higher arousal than neutral images ($M = 3.31$). A list with the selected images, and their average normative valence and arousal is available in the OSF page of the project.

All images were originally colored but changed into grayscale to not interfere with the colors presented in the visual working memory task (as done in Xie and Zhang, 2016).

Participants also completed nine practice trials for which nine separate images were selected, three for each valence condition. Each image was used once during each experimental session, and then repeated in the second session. The colors used as memoranda in each trial were sampled, without replacement, from 360 values evenly distributed along a continuous color circle defined in the CIELAB color model with $L = 70$, centered at $a = 20$, $b = 38$, and radius = 60 (Zhang & Luck, 2008).

Procedure

For all experiments run in Switzerland, the experimental task was programmed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) implemented in Matlab. Participants completed the task in individual booths where they sat approximately 60 cm away from the computer screen (viewing distance unconstrained).

At the start of each trial, an image was presented on the screen with a 9-point Self-Assessment Manikin (SAM) scale below it (see inset in Figure 1A). Values in the SAM ranged from one to nine, with one being very negative and nine being very positive. Participants were instructed to choose the manikin on the left (value = 9) if the presented pictures made them feel happy, pleased, content, hopeful, or anything similar. They should pick the manikin on the right (value = 1), if the image caused feelings of unhappiness, annoyance, discontent, sadness, disgust, and the like. Values in-between indicated intermediate emotional states. Participants were allowed a maximum of 6 s to enter their valence rating of the image. If they did not

respond within this time, a rating of 0 was recorded and the trial continued. Immediately after they rated the image, the screen changed to a fixation cross for 0.8 s and thereafter, an array of five colored dots (radius 1°) was presented for 0.4 s (see Figure 1A). The location of the dots was selected randomly in each trial from eight positions evenly spaced on an imaginary circle (radius = 4.5°) centered in the middle of the screen. Participants were instructed to remember the colors of the dots as precisely as possible. The offset of the memory array was followed by a 1-s blank retention interval. In No-Cue trials, the retention interval ended with presentation of the test display. In Retro-Cue trials, the retention interval ended with presentation of a central white arrow for 0.1 s. The arrow pointed to the location previously occupied by a memory item, thereby indicating with 100% validity that this will be tested. After another 1-s blank interval, the test display was shown. The test display consisted of the color-wheel (randomly rotated from trial-to-trial), a white circle frame around the location of the to-be-recalled dot, and the mouse cursor presented in the center of the screen. Participants were instructed to click on the color in the wheel that best represented the color they remembered seeing in the location marked by the circle frame. All participants completed 60 trials of each possible combination of emotion condition (negative, neutral, positive) and cue condition (no-cue, retro-cue) per session, for a total of 120 trials per condition across the two sessions.

Figure 1

Flow of events in a fictive trial of Experiment 1 (Panel A) and Experiment 3 (Panel C), and an Example Display of the Recall Phase in Experiment 2 (Panel B)

Note. The inset in A shows the details of the 9-point SAM scale. For the colored version of this figure, see the online version of the article. The illustrative image presented here does not belong to the IAPS.

Data analysis

We assessed how self-rated emotion differed between valence conditions. This was taken as indication for the effectiveness of the manipulation of emotion. Next, we assessed performance in the visual working memory task using two measures. First, we computed a model-free index of performance in the task by taking the absolute angular deviation between the response and the actual color of the test item (hereafter referred to as recall error). Second, the distribution of responses in this task was submitted to mixture modeling (Zhang & Luck, 2008). In this model, responses are assumed to consist of a mixture of trials where the target was correctly remembered with a given precision, and trials where participants guessed randomly. Trials with correct memory are modeled as a normal distribution for circular space (i.e., a von Mises) centered on the color of the tested item, and guessing is modeled as a uniform distribution. The model estimates two parameters: the probability that responses were informed by memory as opposed to guessing (probability of recall), and the standard deviation of the distribution reflecting the imprecision of the memory representation (with lower values reflecting better precision).

We fitted the mixture model using the Bayesian hierarchical implementation available in the CatContModel package (Hardman, 2016; Hardman, Vergauwe, & Ricker, 2017) implemented in R (R core team, 2017). All parameter values were determined through Bayesian Markov chain Monte Carlo (MCMC). We fitted the model using 200'000 iterations, and we discarded 100'000 iterations as burn-in to make sure parameter estimation was stable. Hierarchical models assume that the parameter values of individuals in a given condition are drawn from a population-level distribution. Hence in this type of model, we are interested in

the estimate of the population-level parameter in each condition (i.e., the parameter posterior).

All inferences reported here were based on Bayesian hypothesis testing. In Bayesian inference, the parameter estimate is described not only by its mean, but also by the uncertainty around this value (Kruschke, 2011). One can summarize parameter uncertainty by the interval covering 95% of the parameter posterior (i.e., its 95% credible interval). To compare estimates across conditions, we can assess whether posterior differences between conditions include zero, and we can use Bayes Factors (BF). The BF is the ratio of the likelihood of the hypotheses under comparison given the observed data (e.g., the Alternative hypothesis over the Null hypothesis; represented by BF_{10}). $BF_{10} > 3$ are usually regarded as providing substantial evidence for the Alternative hypothesis, whereas $BF_{10} < 0.33$ are considered substantial evidence for the Null hypothesis (Jeffreys, 1961; Wetzels & Wagenmakers, 2012).

Bayesian analyses require that priors on the effects are specified. We used the default priors available in BayesFactor package (Morey & Rouder, 2015) and in the CatContModel package (Hardman et al., 2017). The CatContModel requires two types of priors: priors on the participant's parameters, and priors on the condition effects (i.e., how the parameters vary between conditions). For the individual participant's parameters, there were priors for their mean and variance. For the condition effect on the parameters, the priors were Cauchy distributions with location 0 and a scale value which was set to 2 for the probability of recall parameter, and 5 for the standard deviation parameter. Overall, the priors used were rather uninformative, and they should not have much influence on the estimated group-level posteriors (Rouder, Haaf, & Vandekerckhove, 2018). Hence evaluation of the posterior of the

effect should provide information about the size of the effect in the data relatively independent from the prior beliefs on the size of these effects.

We also fitted the data of each individual participant in each condition using Maximum Likelihood Estimation (MLE) using the simplex algorithm implemented in Matlab (The Mathworks, Natick, MA, US). The results of this analysis for Experiments 1-5 are presented in the Supplementary Materials available at the OSF: <https://osf.io/wzvmx/>. There was perfect agreement between the two types of modeling approaches.

Results

Self-Rated Emotion

Figure 2 and Table 2 present the mean ratings provided for images in each valence condition, and Table 2 presents the evidence for differences in self-rated emotion between conditions. Valence ratings varied monotonically with valence category, being smallest in the negative condition, intermediate in the neutral condition, and largest in the positive condition. Accordingly, there was overwhelming evidence for differences between valence conditions (see Table 2). It is worth noting that the average ratings provided by participants were closely related to the average normative values from the IAPS (see Supplementary Materials on the OSF). Ratings obtained here are also consistent with the ones reported by Xie and Zhang (2016). From this data, we can infer that the images were likely effective in changing participant's self-rated emotional state.

Figure 2*Mean Self-Rated Emotion in Experiments 1 to 4.*

Note. Neg = negative; Neu = neutral; Pos = positive. Small colored dots show individual-level data. Colored triangles represent group-means. Error-bars depict 95% within-subjects confidence intervals.

Table 2

Means and 95% Within-Confidence Intervals for Valence Ratings in each Condition and Evidence (BF_{10}) for Differences between Conditions.

Exp.	Valence Rating			BF_{10} – t -test		
	Negative	Neutral	Positive	Neg vs. Neu	Neu vs. Pos	Neg vs. Pos
1	2.90 [2.64, 3.15]	5.07 [4.97, 5.17]	6.87 [6.59, 7.16]	3.80×10^{12}	9.01×10^8	5.92×10^{11}
2a	2.59 [2.40, 2.78]	5.16 [4.99, 5.34]	7.31 [7.01, 7.61]	3.97×10^{14}	4.9×10^7	1.17×10^{13}
2b	2.71 [2.50, 2.92]	5.11 [4.94, 5.28]	7.25 [7.00, 7.50]	7.1×10^{11}	1.28×10^9	3.29×10^{13}
3	2.55 [2.35, 2.76]	5.06 [4.99, 5.13]	7.05 [6.84, 7.26]	4.99×10^{24}	2.67×10^{20}	4.73×10^{24}
4	2.23 [1.99, 2.46]	5.05 [4.87, 5.22]	7.46 [7.12, 7.80]	3.24×10^{12}	1.41×10^7	1.06×10^{11}

Note. Neg = negative; Neu = neutral; Pos = positive.

Recall Error

Figure 3 presents the mean recall error in each experimental condition, and Table 3 presents the evidence for differences in recall error between valence conditions as assessed with Bayesian t -tests. As shown in Figure 3, there was no difference in average recall error between valence conditions. In contrast, the attentional manipulation was effective: recall error was substantially smaller in the retro-cue condition compared to the no-cue condition. We submitted the recall error to a Bayesian ANOVA having valence and retro-cueing as predictors. The best model of the data included only the main effect of retro-cueing ($BF_{10} = 2.53 \times 10^{40}$). There was substantial evidence against including the main effect of valence in the best model ($BF_{10} = 0.12$), and of an interaction between valence and retro-cueing ($BF_{10} = 0.11$).

Figure 3*Recall Error in Experiments 1, 2, 3, and 4*

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Small colored dots present individual-level means in each condition. Triangles depict the group-average in each condition. Error-bars represent 95% within-subjects confidence intervals.

Table 3*Evidence (BF_{10}) for Differences between Valence Conditions in Recall Error.*

Exp.	Other Factor	BF_{10} – t-test (two-sided)		
		Neg vs. Neu	Neu vs. Pos	Neg vs. Pos
1	No-Cue	0.26	0.53	0.26
1	Retro-Cue	0.23	0.32	0.32
2a		0.24	0.25	0.26
2b		0.28	3.29	3.45
3		0.29	0.17	0.21
4	RI = 1 s	0.25	0.31	0.43
4	RI = 3 s	0.26	0.26	0.28

Note. RI = retention interval. Neg = negative; Neu = neutral; Pos = positive. $BF > 3$ are printed in bold.

Mixture Modeling

We modeled the data to obtain parameters reflecting the quality of the memory representation (aka memory imprecision) and the quantity of stored representations (aka probability of recall). Figure 4 presents the group-level estimates of memory imprecision and probability of recall in each experimental condition.¹ Table 4 presents the evidence for the main effects of valence and retro-cueing, and their interaction.

Valence had no effect on estimates of memory precision or probability of recall (i.e., the evidence favored models without an effect of valence, see Table 4). Retro-cues, in contrast, increased memory precision and the probability of recalling the tested item. There was strong

¹ The Supplementary Materials File available in the OSF (Section B) shows posterior predictive checks indicating how well the model captured the distribution of errors in each condition across all experiments reported here. Posterior predictive consists of simulated data based on the likely model parameters.

evidence against an interaction between valence and retro-cue. Of particular interest is the contrast between the negative and the neutral conditions, which was the critical finding in the study of Xie and Zhang (2016). There was ambiguous to substantial evidence *against* differences between the negative and neutral conditions in the no-cue condition ($BF_{10} = 0.39$) and retro-cue ($BF_{10} = 0.23$) conditions. But if anything, memory precision was slightly *higher* in the neutral than the negative condition.

Figure 4

Group-Level Estimates for the Mixture Model Parameters in Experiment 1

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Error-bars present the 95% credibility interval of the parameter posterior.

Table 4

Bayes Factors in Favor of the Alternative Hypothesis (BF_{10}) for Effects of Manipulated Variables on Parameters of the Mixture Model.

Exp.	Factor	Model Parameter	
		Memory Imprecision	Probability Recall
1	Valence	0.012	0.006
	Retro-Cue	1489	1.29×10^{37}
	Valence x Retro-Cue	0.010	0.001
2a	Valence	0.065	0.010
2b	Valence	0.090	0.112
3	Valence	0.008	0.002
4	Valence	0.23	0.004
	Retention	5.03	0.187
	Valence x Retention	0.536	0.013

Note. BFs > 3 are printed in bold.

Discussion

We presented participants with a dual-task scenario: first, they saw an image and they rated its valence; second, they had to study a set of colored dots for a color reproduction test. This procedure was closely modeled after the study of Xie and Zhang (2016, Experiment 1). The images were selected from a standardized database (the IAPS) to elicit negative, neutral, and positive emotions. Accordingly, valence ratings in Experiment 1 closely matched the normative ratings used to select the images: participants consistently reported that negative, neutral, and positive images elicited negative, neutral, and positive emotions on them, respectively. Despite successfully manipulating the self-reported emotional state of the participants, we did not observe an effect of self-reported emotion on visual working memory. Model-free indicators of performance (recall error), and estimates of memory precision and probability of recall obtained from a hierarchical implementation of the mixture model both provided evidence

against an effect of self-rated valence and, in particular, of self-rated negative emotion. In other words, we failed to replicate the results of Xie and Zhang (2016).

Performance in our study was not atypical. Experiment 1 replicated the well-known effect of retro-cueing: memory precision and probability of recall increased when participants focused their attention on the to-be-tested item (e.g., Murray, Nobre, Clark, Cravo, & Stokes, 2013; Pertzov, Bays, Joseph, & Husain, 2013; Souza, Rerko, Lin, & Oberauer, 2014; Souza et al., 2016).

At this point, we did not know what could have led to the absence of an effect of self-rated emotion in our study. Our sample size ($N = 22$) was similar to that of Xie and Zhang (2016; $N = 18$) and the number of trials per design cell was identical ($n = 120$). There were, however, some procedural differences. First, and most obvious, we included a very powerful manipulation of attention in half of the trials, and we conducted two sessions to reach the same number of trials per design cell. Arguably, the inclusion of this manipulation could have hindered our chances of observing a negative emotion boost for various reasons. Second, we only cued the tested-item and left the test display otherwise empty, whereas Xie and Zhang (2016, Experiment 1) cued one item to be recalled, while at the same time the other non-tested items were presented onscreen. Presentation of the non-tested items reduces the possibility for misbinding errors (i.e., recall of another item than the tested one), thereby arguably improving performance. Third, we noticed that we allowed participants more time to rate the image (max. 6 s) than in Xie and Zhang (4 s). In Xie and Zhang (2016), the picture remained onscreen for 4 s irrespectively of how quickly participants rated the image. In Experiment 1, the image rating task was immediately removed after the rating was entered. Our reasoning for

presenting the working memory trials immediately after the rating was to prevent the emotion from dissipating after the image was consciously processed. Any of these differences could potentially explain the lack of an effect of emotion in our study. Experiment 2 was designed to address these concerns.

Experiment 2

Experiment 2 was designed to replicate the study of Xie and Zhang (2016) more closely. First, we removed the manipulation of attention. All working memory trials only involved a blank screen during the retention interval. Second, we presented the non-tested items in the test display thereby potentially reducing the probability of misbinding errors. Regarding differences in the rating task, we implemented two experimental versions. In Experiment 2a, the rating task was exactly as implemented in Experiment 1. In Experiment 2b, the rating task lasted 4 s, exactly as in Xie and Zhang (2016).

Method

A new sample of twenty students (16 women; $M = 22.45$ years old) from the University of Zurich participated in Experiment 2a, and another sample of 20 students (17 women, $M = 24.3$ years old) took part in Experiment 2b. As indicated in the methods section of Experiment 1, this sample size should be sufficient to detect an effect as large as the one reported by Xie and Zhang (2016) with a power of 80%. Both experiments consisted of a single 1-hour session and participants received 15 CHF as reimbursement.

The same task as used in Experiment 1 was employed in Experiment 2 with the following exceptions. First, we modified the test display to present one empty circle-frame (indicating the

test item) and four colored dots (the non-tested memory items) as illustrated in Figure 1B. This procedure limits the possibility that participants would retrieve the incorrect item from memory, and it better aligns with the procedure employed by Xie and Zhang (2016). Second, we removed the retro-cue trials from the design, leaving only the standard color-reproduction trials that are typical in the visual working memory literature. Third, we reduced the number of images in each category to 60, which were used twice during the session thereby closely matching the procedure used by Xie and Zhang (2016). We retained in the experiments the 60 images² that were, in average, rated closest to the normative ratings used to select them (see Supplementary Materials in the OSF). This was done to ensure that we maintained images that participants mostly agreed upon with regards to the emotion they elicited. The average normative valence ratings of the retained images were: (a) negative, valence $M = 2.69$, arousal $M = 5.2$; (b) neutral, valence $M = 4.99$, arousal $M = 3.2$; and (c) positive valence $M = 7.22$, arousal $M = 4.87$.

Fourth, whereas in Experiment 2a the rating task lasted maximally 6 s and was finished as soon as participants entered their ratings (as done in Experiment 1), in Experiment 2b we reduced the time available for rating the image to 4 s, and we maintained the image onscreen for 4 s irrespective of how long it took for participants to rate them, thereby equating conditions with regards to exposition to the images. To make sure participants were not confused about whether their response was registered, the valence scale and the mouse cursor disappeared after the participant made their rating, and only the image remained onscreen for the remainder of the 4 s. Fifth, in Experiment 2b we included breaks every 40 trials in which a

² A list of the retained images is available on: <https://osf.io/wzvmx/>.

message reminded participants that they should rate the image as precisely as possible with regards to how it made them feel.³

Results

Self-Rated Emotion

Average self-rated valence in Experiments 2a and 2b were similar to the ones obtained in Experiment 1 (see Figure 2 and Table 2), and there was overwhelming evidence for differences in ratings between the three valence conditions.⁴ In sum, in both experiments, self-reports indicated that negative, neutral, and positive images elicited the corresponding emotions on the participants.

Recall Error

Recall error in each condition is presented in Figure 3, and evidence for differences between valence conditions in this measure is presented in Table 3. There was substantial evidence against an effect of valence in Experiment 2a. In Experiment 2b, there was substantial evidence for lower recall error in the positive valence condition compared to either the neutral or negative valence conditions.

³ In personal communication with Dr. Xie, he clarified to us that an RA was present throughout the experimental session and reminded participants of this instruction every 40 trials. In our lab, the experimenter does not interact with the participants during the task, so we included a printed message to remind participants of the instructions.

⁴ Supplementary Materials on the OSF shows that ratings of each image were similar to the ones in the IAPS norms.

Figure 5*Group-Level Estimates for the Mixture Model Parameters in Experiment 2*

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Error-bars present the 95% credibility interval of the parameter posterior.

Mixture Modeling

Figure 5 presents the group-level parameter estimates in Experiment 2 and Table 4 presents evidence for the main effect of valence. For both experimental versions, there was no evidence for a main effect of valence in either mixture model parameter. Again, if anything, estimates of memory precision were *smaller* in the neutral condition compared to the negative one.

Discussion

In Experiments 2a and 2b, we systematically reduced differences between our experimental set-up and the one employed in the Experiment 1 reported by Xie and Zhang (2016). Again, participants in our Experiment 2 rated the IAPS images in accordance with their normative valence ratings, showing the expected modulations of their emotional state. Nevertheless, we could not find evidence for differences in visual working memory performance across valence conditions. In particular, we again saw no evidence that self-rated negative emotion increased visual working memory precision.

There were no differences left between our experimental set-up and the one of Xie and Zhang (2016). A close reading of their paper, however, indicated that the evidence for differences in memory precision between the negative and neutral conditions in Xie and Zhang (2016, Experiment 1) was ambiguous: the Bayes factor for the comparison between these conditions was only $BF_{10} = 2.2$. Xie and Zhang (2016), however, had a second experiment using shapes as the memoranda and colored IAPS images for rating that yielded stronger evidence for an effect of self-rated negative emotion on visual working memory precision ($BF_{10} = 6.5$). Perhaps the use of grayscale images and colored dots as memoranda was not optimal to induce or measure the effects of self-rated negative emotion. We therefore ran Experiment 3 using colored IAPS images and orientation as the memoranda to assess whether this task could yield evidence for a fidelity boost following self-rated negative emotion.

Experiment 3

Experiment 3 involved a conceptual replication of Experiment 2 reported by Xie and Zhang (2016). Xie and Zhang employed colored IAPS images and closed-contour shapes as memoranda in their Experiment 2, and they obtained stronger evidence than in their Experiment 1 for a modulation of memory precision following self-rated negative emotion in this study. We reasoned that the use of colored images to induce emotion and abstract visual stimuli to assess memory could provide a more favorable condition to observe the effects of self-rated negative emotion on visual working memory. There is some evidence that negative images attract more attention and are rated more negatively compared to neutral images when presented in color than in grayscale (Bekhtereva & Müller, 2017). Furthermore, having color as the memoranda might compete with the emotional response generated by the images as colors can also evoke emotional responses/associations (Valdez & Mehrabian, 1994). Therefore, we decided to run a replication attempt using colored emotional images to induce emotion and non-colored visual stimuli as the memoranda. Instead of the closed-contour shapes, we decided to use a continuous orientation reproduction task using isosceles triangles (Fougnie, Asplund, & Marois, 2010). We thought the use of this task was more advantageous because: (a) orientation is a more commonly used task in visual working memory research than the closed-shape contour task (see also Long et al., 2020 for the use of an orientation task); and (b) orientations vary continuously along 360°, whereas the shape-wheel used by Xie and Zhang was just an approximation of a continuous feature space. Their shape space had only 180 shapes, from which only 30 were presented on the test display (but see Zhang & Luck, 2011 for a validation of this procedure).

If there is an effect of negative emotion on memory precision, the set-up of Experiment 3 would be more suitable to measure it. We also made sure this experiment had a larger sample size to have enough power to measure any effects of negative emotion, if they existed.

Method

Xie and Zhang (2016) obtained a $d_z = 0.71$ in their Experiment 2 with a sample of $N = 19$ participants. Given that our previous experiments provided evidence against an effect of self-rated negative emotion, in Experiment 3 we aimed to have a sample size at least twice as large as in the original study to increase power. A power analysis performed in G*Power indicated that for a $d_z = 0.71$, a sample size $N = 14$ yields power $1 - \beta = 80\%$ ($\alpha = 0.05$). A new sample of 42 students (35 women; mean age = 22.3 years) of the University of Zurich took part in Experiment 3 for an 1-hour session in exchange of 15 Swiss francs or course credit.

The experimental task is illustrated in Figure 1C. A colored image drawn from the same pool of images as used in Experiment 2b was presented for 4 s. During this time, participants rated the valence of the image using the 9-point SAM scale. Unlike Experiments 1-2, the IAPS images were presented in color. Thereafter, the fixation cross was shown for 0.8 s, after which an array of four oriented white triangles was presented for study for 0.4 s. The location of the triangles was selected from eight locations evenly spaced on an imaginary circle centered in the middle of the screen. After a retention interval of 1 s, the test display was shown which consisted of a randomly rotated probe triangle and the mouse cursor. Participants adjusted the orientation of the probe triangle using the mouse to match the one they remembered having seen at the probe's location in the memory array. Once they were satisfied with the orientation of the probe, they clicked with a mouse button to confirm their response.

We did not present the non-target items in the test display because Xie and Zhang (2016) also did not present the non-target shapes on the test display in their Experiment 2. Combined with the fact that this did not play a role in our color experiments (Experiment 2 in this paper), this is probably not a critical feature to observe the boosting effect of self-rated negative emotion on visual working memory.

As in the previous experiments, participants completed 120 analyzed trials per valence condition. There were 9 additional practice trials that were excluded from analysis.

Results

Self-Rated Emotion

Figure 2 and Table 2 presents the results of the image rating task. Valence ratings were again commensurate with the normative valence of the images⁵, and all three valence conditions differed substantially from each other. The valence ratings were comparable to the ones obtained for the grayscale images in Experiments 1 and 2, suggesting that color did not change ratings. All in all, these results indicate that the emotion manipulation was again successful.

Recall Error

Figure 3 and Table 3 present the results regarding the recall error measure. Replicating Experiments 1 and 2, there was no evidence for differences between valence conditions.

⁵ See also Supplementary Materials on the OSF.

Mixture Modeling

Figure 6 presents the group-level parameter estimates of the mixture model in Experiment 3. Evidence for the main effect of valence condition is presented in Table 4.

Figure 6

Group-Level Estimates for the Mixture Model Parameters in Experiment 3

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Error-bars present the 95% credibility interval of the parameter posterior.

Mean parameter estimates were similar across valence conditions, and there was overwhelming evidence against a difference in memory precision and probability of recall across conditions in Experiment 3. These results indicate that our failure to observe an effect of self-rated emotion on visual working memory across Experiments 1, 2a, and 2b is unlikely due to the use of grayscale images or having colors as the memoranda.

Discussion

Experiment 3 replicated the design of Xie and Zhang (2016, Experiment 2) with a visual working memory task involving the memorization of the orientation of white triangles, and in which we presented the IAPS images in color. Despite these modifications, our results still confirmed the absence of any effect of self-rated emotion on visual working memory performance.

Experiment 4

While we were running Experiments 1-3 in Switzerland, H. R. Liesefeld independently ran a study in Germany trying to replicate and extend the results of Xie and Zhang (2016). The goal of this study was to assess whether the effects of emotional manipulation would be retained when the retention interval was prolonged (1 s vs. 3 s). This study provides one further attempt by an independent researcher to replicate the boosting effect of negative emotion on visual working memory.

Method

A sample of 17 students (15 women; $M = 24.06$ years old) from Ludwig-Maximilians-Universität München in Germany received course credit for participation in Experiment 4. Sample-size determination was based on obtaining a similar sample-size as in Xie and Zhang (2016). Participants reported normal color vision and visual acuity.

The task was the same as used in Experiment 1 with the following exceptions. First, the stimuli consisted of squares (as in Xie and Zhang, 2016; rather than dots as used in Experiments 1 and 2) and the IAPS images were the same as reported by Xie and Zhang in their

supplementary table. Second, squares were sized 2° and images 11° of visual angle in width as seen from a distance of 70 cm on a TFT screen (1920×1080 pixels; $53 \times 29,8$ cm; 60 Hz). Squares were placed randomly (with a minimal distance between each pair of squares of 45° center-to-center on the circle) on an imaginary circle of 4.5° radius around fixation. Third, to-be-memorized colors were sampled randomly from a slightly different circle in the CIELAB space ($L = 63$, centered on $a = 9$, $b = 27$, and radius = 40). Fourth, the retention interval was varied (balanced and randomly intermixed) with two-levels: 1 s and 3 s. Fifth, for the emotion rating at the beginning of each trial, the image was presented for a fixed duration of 3 s during which participants could respond and only 5 response alternatives were offered (i.e., removing the intermediate steps between the manikins). In the analyses, these five options were assigned the same numeric values (namely, 1, 3, 5, 7, and 9) as in the experiments that offered nine options to facilitate comparison across experiments. Lastly, all squares were filled with their original color in the test display except for the probed item, which was indicated by a black outline frame (similarly to Experiments 2a, 2b, and Xie and Zhang, 2016).

Participants received immediate performance feedback on each trial via a small bar on the color wheel that indicated the correct response and, by implication, how far off the actual response was from the correct color. They performed 12 practice trials followed by 10 blocks of 60 trials each, resulting in 100 trials per Valence $\times$ Retention interval condition. The deviations in details of the experiment compared to the experiments reported above are because the experiments were conceived and run independently.

Results

Self-Rated Emotion

Figure 2 and Table 2 present the results of the image rating task. Self-rated emotion was commensurate with the valence category of the image and similar to the ones obtained in Experiments 1-3.

Recall Error

Figure 3 and Table 3 present the evidence for differences between valence conditions across the two levels of retention interval. There was substantial evidence *against* differences between valence conditions in terms of recall error, except for one comparison (negative vs. positive in the 1-s retention interval), which was in the inconclusive range of BF_s.

Mixture Modeling

Figure 7 presents the group-level estimates of the mixture model in each experimental condition. The overall analysis indicates that there was no evidence for a main effect of valence on either memory precision or probability of recall. Retention interval, however, had some effect on memory precision, but not on the probability of recall.

Inspection of Figure 7 indicates that memory precision was higher in the negative and positive valence condition compared to the neutral one when the retention interval was 1 s. This somewhat higher memory precision is in line with an effect of arousal (because both negative and positive images were more arousing). That the effect vanishes when the retention interval was longer might indicate that the effect of arousal dissipates across time and affects recall rather than encoding or maintenance. However, given that this effect is relatively weak

and that there was no effect of arousal in any of the other studies, we would discourage putting much trust in it. Rather, combined with the fact that the effect of arousal was not supported by the ANOVA, nor observed for the model-free index, these results suggest again that self-rated emotion does not substantially change processing of visual information in working memory.

Figure 7

Group-Level Estimates for the Mixture Model Parameters in Experiment 4

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Error-bars present the 95% credibility interval of the parameter posterior.

Discussion

Again, we observed modulations of self-rated emotion as a function of presentation of the images. Yet again, we did not obtain evidence that self-rated negative emotion selectively increased visual working memory precision. In one condition (1-s retention interval), memory precision was estimated to be somewhat higher after self-rated negative emotion than after

self-rated neutral emotion. But the same pattern was observed for the self-rated positive valence condition. These results are generally inconsistent with the proposal that self-rated negative emotion changes processing style towards more detail. If anything, they are consistent with the possibility that high arousing information can improve storage of visual details. Given that this pattern was only observed in one condition, and one experiment, we would refrain from interpreting it as reflecting a reliable pattern. Rather, this finding highlights to us that individual experiments can yield variable results, which becomes an issue when these are selectively published (Rosenthal, 1979).

Experiment 5

Experiment 5 was inspired by the study of Spachtholz et al (2014). In Spachtholz et al. (2014), negative or neutral emotion was induced with an autobiographical task (and presentation of emotion-congruent music) in two separate groups of participants. Thereafter, participants completed a visual working memory task similar to the one employed in our color visual working memory experiments (Experiments 1, 2, and 4). Participants also completed trials in which the retention interval was set to 0, in which case they could rely on their iconic memory (hereafter the iconic memory task). For both tasks, the authors reported that negative emotion reduced the quantity, but increased the quality of the visual representations.

Experiment 5 used a similar logic. We exposed groups of participants to either a negative, neutral, or positive mood induction phase where negative, neutral, or positive images from the IAPS together with emotion-congruent music were presented for 3 min. Next, participants completed the continuous color reproduction task. In a separate block, participants

completed the continuous color reproduction task without emotional induction to assess their baseline level of performance.

Given that our original interest was on assessing changes in attentional processing within working memory induced by emotions, this study also included a manipulation of retro-cueing. In half of the trials, a cue indicated which item was relevant for the test 1-s before the test situation was presented.

Method

Participants

Our aim was to collect data of 20 participants in each valence group (between-subjects). Spachholz et al. (2014) reported a $z = -2.35$ for a sample of $N = 80$ participants in a between-subjects design, which corresponds to an effect-size of $d = 0.55$. Given that the design of this experiment involved comparing performance of the participants in each group before and after emotion induction, the same power analysis as done for Experiment 1 applies here: this sample can give a power of 80% to detect an effect of $d = 0.58$.

Participants were 126 undergraduate students from University of Minho, Portugal. Participants received extra-course credit as reimbursement for taking part on the study. Sixty-seven participants were excluded from the final data analysis, due to: computer malfunction during the procedure implementation ($n = 2$, both originally in the neutral group); depressive and anxiety symptoms when evaluated through Beck's Depression Inventory and Beck's Anxiety Inventory ($n = 25$; 8 from the negative group, 8 from the neutral, and 9 from the positive group); and inconsistency between the applied emotional induction and the reported emotion

($n = 39$; 20 from the neutral group, 14 from the negative group, and 5 from the positive group).

The Supplementary Materials provide a list detailing which emotion was reported by the excluded participants split by group and exclusion reason (except computer mal-function).

These results show that the neutral group tended to vary more on their reported emotions.

Excluded participants in the negative group tended to report neutral feelings, indicating that the negative induction was perhaps too mild to reduce the general positive mood in which most participants seemed to be (as indicated by the higher success of the positive condition and their PANAS ratings presented later).

The final sample consisted of 60 participants (37 women, $M = 21.4$ years old), which were evenly assigned to the Negative ($n = 20$), Neutral ($n = 20$), and Positive ($n = 20$) induction groups. All participants completed an informed consent form in the beginning of the experiment and were debriefed afterwards. The experimental protocol was approved by the Ethics Committee of the University of Minho (SECSH 009/2014).

Instruments

At the end of the experiment, participants completed the Portuguese version of the Beck's Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Vaz Serra & Pio Abreu, 1973), and the Portuguese version of the Beck's Anxiety Inventory (Beck, Epstein, Brown, & Steer, 1988; Quintão, Delgado, & Prieto, 2013). Before and after the experiment, participants also completed the short-form Portuguese version of the Positive and Negative Affect Scale (Galinha & Pais-Ribeiro, 2005; Galinha, Pereira, & Esteves, 2014; Watson, Clark, & Tellegen, 1988).

Emotional Induction Procedure

Each participant was exposed to only one type of emotional induction (Negative, Neutral, or Positive group). Emotional induction was conducted through the simultaneous presentation of visual and auditory stimuli for a total duration of 3 min.

Ninety images were selected from the IAPS database adapted for Portugal (Soares et al., 2015): 30 negative images ($M = 2.99$; $SD = 0.14$), 30 neutral images ($M = 5.20$; $SD = 0.20$), and 30 positive images ($M = 7.05$; $SD = 0.14$).⁶ Auditory stimuli consisted of classical musical pieces chosen based on the review by Västfjäll (2001). The selected songs were edited in the Audacity® program to be only three minutes in length, and fade-in and fade-out phases were inserted where appropriate. The song presented in the Negative induction phase was Albinoni's "*Adagio*", with a fade out of 0.99 s at the end of the excerpt. Bach's "*Brandenburg Concert N.2*" was presented in the Positive induction phase. We included a fade-in of 1.20 s at the beginning of the song and a fade out of 6.62 s at the end. In the Neutral induction phase, Holst's song "*The Neptune - The Mystic*" was played, with an added fade-in and fade-out of 3 seconds at the beginning and the end of the song, respectively.

During the induction procedure, participants were presented with one image at a time (duration 3 s), while music was played through headphones (see Figure 8A). Participants were instructed to look carefully at the images, to listen to the music, and to focus on how the images and music made them feel. At the end of the induction, nine adjectives were randomly placed onscreen on a 3 × 3 grid. Three adjectives were associated with each valence condition (here roughly translated from Portuguese): "*melancholic*" (melancólico), "*sad*" (triste) and

⁶ The list of selected images is available at the OSF page of the article.

"*distressed*" (angustiado) were assumed to reflect a negative mood; "*indifferent*" (indiferente), "*neutral*" (neutro), and "*alien*" (alheio) were assumed to reflect a neutral mood; and, "*happy*" (feliz), "*enthusiastic*" (entusiasmado), and "*euphoric*" (eufórico) a positive mood. Participants were instructed to select one of these adjectives to characterize how they felt while seeing and hearing the stimuli presented (valence measure). Next, they reported on a scale from 1 to 7, the intensity with which they felt the emotion described by the adjective selected (arousal measure).

The emotional induction procedure was considered effective if: (a) there was coherence between the emotion reported by the participant and the induced emotion (e.g., a negatively induced participant should choose one of the three negative adjectives to characterize their state); and (b) the reported emotion intensity was greater than three.

Visual working memory task

The visual working memory task consisted of a color reproduction task similar to the one employed in Experiment 1 (see Figure 8B). At the beginning of each trial, a white fixation cross appeared in the screen center for 0.5 s. Thereafter, the memory array consisting of six colored discs (radius = 32 pixels) were shown evenly spaced on an imaginary circle (radius = 150 pixels). The memory array was presented for 1 s simultaneously with the color wheel (defined as in Experiment 1). Then a retention interval followed for 1 s in which only the gray background and the color wheel were visible onscreen. In no-cue trials, this interval ended with the presentation of a recall cue (a white circle frame in the location of one of the memory items) and the mouse cursor. In retro-cue trials, the recall-cue was presented onscreen for 1 s at the same time as the color wheel was removed from the screen. This manipulation served to

guide attention to the to-be-retrieved information. After this delay, the color wheel reappeared together with the mouse cursor (test screen). Participants were instructed to select the color of the item indicated by the recall cue as precisely as possible. Furthermore, to prevent verbal labelling of the colors (Souza & Skóra, 2017), participants were asked to repeat “BA-BI-BU” aloud throughout the trial.

Figure 8

Induction Procedure and Visual Working Memory Task in Experiment 5

Note. The images used here are not part of the IAPS. These are royalty-free images from unsplash (<https://unsplash.com/>).

Procedure

Participants completed the visual working memory task in two blocks: one without and one with emotional induction. Each block consisted of 150 trials which were evenly and randomly divided into retro-cue and no-cue trials. The order of the induction and no-induction blocks was counterbalanced across participants.

In the no-induction/induction experimental counterbalancing order, participants started the experiment by completing the PANAS. Then they completed 150 trials of the visual working memory task (no-induction block). At the end of this block, they left the experimental room for 10 min and completed an unrelated distractor task. Then they returned to the experimental room, received the induction procedure for 3 min, reported their emotional state, and completed the second block of visual working memory trials (induction block). Lastly, they completed the PANAS.

In the induction/no-induction counterbalancing order, after completing the PANAS, participants were immediately exposed to the induction procedure, reported their emotional state, and completed the first block of visual working memory trials (induction block). When this block was over, they left the room for 10 min to complete a distractor task. When they returned, they completed the no-induction block of the visual working memory task.

Regardless of the order in which they performed the tasks, in the end all participants completed the Beck's Depression Inventory and Beck's Anxiety Inventory, which were used to exclude from the analysis the subjects who showed symptoms of any psychological distress. Finally, participants were debriefed regarding the purpose of the study.

Results

Panas

In the beginning and the end of the session, participants completed the short-form Portuguese version of the PANAS. Figure 9 presents the ratings for the positive and negative affect sub-scales of the PANAS (range: 1 – 25) at the start and end of the session across groups. Participants reported high levels of positive affect and low levels of negative affect at the start of the experiment. At the end of the experiment, both ratings tended to decrease (with this decrease being larger for positive ratings). In general, valence groups did not differ in their levels of positive and negative affect at the start and end of the session.

A Bayesian ANOVA having completion time (start vs. end), affect scale (positive vs. negative) and group (negative, neutral, and positive) as predictors showed that the best model of the data included the main effects of affect scale, completion, and their interaction ($BF = 2.05 \times 10^{63}$). There was strong evidence against including an effect of group in this model ($BF = 0.079$). There was overwhelming evidence for the main effects of affect scale (reflecting the higher ratings for positive than negative affect) and for completion time (reflecting lower ratings at the end of the experiment), inclusion $BFs > 1000$, and strong evidence for the interaction, inclusion $BF = 112$, reflecting the fact that positive ratings decreased more from the start to the end of the session.

Figure 9*Mood Ratings at the Start and End of Experiment 5*

Note. Pos. = positive. Neg. = negative. Summed score for each individual (max. = 25) is presented as small dots. The group average is displayed as a triangle. Error-bars represent 95% within-subjects confidence intervals. Large values indicate higher levels of the respective affect.

Emotion Rating Post Induction

In the Negative group, the adjectives "*melancholic*", "*sad*", and "*distressed*" were selected by 6, 4, and 10 participants, respectively. In the Neutral group, the adjectives "*indifferent*", "*neutral*", and "*oblivious*" were selected by 5, 11, and 4 participants, respectively. In the Positive group, the adjectives "*happy*", "*enthusiastic*" and "*euphoric*" were selected by 14, 6, and 0 participants, respectively. The average intensity (on a scale from 1-7) of the reported emotion was 4.70 ($SD = 0.80$), 4.74 ($SD = 1.19$), and 4.75 ($SD = 0.91$) in the Negative, Neutral⁷, and Positive groups, respectively.

⁷ Intensity rating for one participant in this group was lost due to an error.

Recall Error

Figure 10 presents the recall error across conditions in Experiment 5. We submitted this data to a Bayesian ANOVA having valence group (negative vs. neutral vs. positive; between-subjects), retro-cueing (cue vs. no cue; within-subjects), and emotion induction (with vs. without induction; within-subjects) as predictors. The best model of the data included only the main effect of retro-cueing ($BF = 25.5$), reflecting that performance was credibly better in the retro-cue condition compared to the no-cue condition.

There was ambiguous evidence against including in the best model the effect of valence group ($BF = 0.52$) and against including the effect of emotion induction ($BF = 0.36$). Critically ⁸, there was evidence against including the interaction between induction and valence group ($BF = 0.16$), indicating that emotion induction did not change performance in the working memory task. There was also strong evidence against an interaction between valence group and retro-cue ($BF = 0.09$) and between retro-cue and induction ($BF = 0.25$). The evidence against inclusion of the three-way interaction was ambiguous ($BF = 0.67$).

⁸ To isolate evidence for the interactions, we contrasted models including the main effects of interest only vs. models which in addition included the interaction of interest.

Figure 10*Recall Error in Experiment 5*

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Small colored dots present individual means in each condition. Triangles depict the group-average in each condition. Error-bars represent 95% within-subjects confidence intervals.

Mixture Modeling

Given that Experiment 5 involved a between-subject design, we first fitted the data of each group separately (number of iterations = 200'000, burn-in = 75000), and then we pulled together the posteriors of all groups to assess for the combined evidence for the effects of the between-subjects variables and the within-subjects variables. Figure 11 presents the posteriors of the memory imprecision and probability of recall estimates in each condition. Table 5

presents the evidence for the effects of the manipulated variables on the parameters of the mixture model.

Overall, there was substantial evidence against an effect of induction and its interaction with valence group on any parameter of the mixture model. The only effect that received some support was of retro-cueing on the probability of recall, because participants were somewhat more likely to retrieve the tested item in the retro-cue condition compared to the no-cue condition.

Figure 11

Group-Level Estimates for the Mixture Model Parameters in Experiment 5

Note. Neg = Negative; Neu = Neutral; Pos = Positive. Error-bars present the 95% credibility interval of the parameter posterior.

Table 5*Evidence for the Effects of Valence Group, Induction, and Retro-Cueing in the Mixture Model**Parameters of Experiment 5.*

Factor	BF ₁₀	
	Imprecision	Prob. Recall
Valence Group	0.049	0.015
Induction	0.186	0.045
Retro-Cueing	0.109	3.953
Valence Group x Induction	0.028	0.019
Valence Group x Retro-Cueing	0.065	0.001
Induction x Retro-Cueing	0.033	0.004
3-way	0.023	0.020

Note. BFs > 3 are printed in bold.**Discussion**

Experiment 5 examined the question whether inducing one emotional state (either a negative, neutral, or positive state) would change performance in the continuous visual working memory task, trying to replicate the findings of Spachtholz et al. (2014). Furthermore, one of our goals was to assess whether this emotional induction would change the way attention is engaged for retrieving information from working memory (i.e., the retro-cue effect).

To assess the impact of emotion induction, we sampled three groups of participants. We measured their visual working memory performance without emotional induction and after emotional induction, allowing us to isolate changes in performance due to the induced emotion while controlling for individual differences in working memory performance. We only retained in the final sample, individuals that reported feeling the corresponding emotion after the induction procedure to make sure the induction was effective in changing participant's emotional state. We note, however, that given the high rate of unsuccessful induction within

the neutral and negative groups, the type of emotional induction used in this experiment cannot be considered particularly strong or effective. Nevertheless, when we considered only participants that reported negative emotion in line with the induced state, we still did not find evidence of a change in performance in the visual working task in comparison to their own performance without the emotional induction and the other valence groups.

The only predictor that affected performance in Experiment 5 was the manipulation of retro-cueing. In Experiment 5, the retro-cue effect was operationalized as the presentation of the recall cue while at the same time removing the opportunity for participants to respond in the task (i.e., the color wheel). This forces participants to use the cue only to retrieve information from working memory. The benefit observed in Experiment 5 replicates previous studies (Souza, Rerko, Lin, et al., 2014; Souza et al., 2016) showing that retrieval can be improved if participants retrieve the relevant information ahead of testing (see Souza & Oberauer, 2016 for a review). Crucially, the manipulation of emotion did not modulate the retro-cue effect.

One concern in Experiment 5 refers to the number of trials. We had a lower number of trials in each design cell (75) compared to the remaining experiments in this series (E1, E2, and E3 = 120, whereas E4 = 100), which was also smaller than in the published studies (Xie & Zhang = 120; Spachtholz et al = 96). The main reason for using a high number of trials in studies of this type is to gain sufficient information on the distribution of recall errors to derive reasonably accurate parameter estimates from the mixture model. However, simulation studies have shown that hierarchical Bayesian models (as employed here) can accurately recover the parameters of the mixture model even with a low number of trials (Oberauer, Stoneking,

Wabersich, & Lin, 2017). Indeed, our hierarchical mixture modeling provided an adequate fit to the data (see Supplementary Materials in the OSF). Hence, given the previous experiments in this series, it is more likely that the absence of valence effects observed here reflects a lack of effect of self-rated emotional states on visual working memory performance.

Experiment 6

So far, our experiments have provided evidence against a precision boost induced by self-rated negative emotion. There are, however, three critiques that can be applied to our experiments. The first one refers to cultural and socio-economic differences: samples in Experiments 1-5 consisted of university students from three European countries, with most experiments having been conducted in Switzerland for which the socio-economic status is likely higher than for the North American samples in the study of Xie and Zhang (2016). A recent meta-analysis indicated that the effect of emotional induction is somewhat larger among North American samples compared to European ones (Joseph et al., 2020). Second, our sample sizes were relatively modest – ranging from 17 to 42 participants. These sample sizes were similar to the one reported by Xie and Zhang (2016). Nevertheless, if the effect of self-rated emotion is much smaller than estimated from the original study, our experiments would be highly underpowered to detect it. Third, we have contrasted the effect of valence across three levels: negative, neutral, and positive. According to the published studies, however, the critical contrast is between self-rated negative vs. neutral emotion. Positive emotion is generally more difficult to induce than negative emotion (Joseph et al., 2020). Hence a more focused contrast

between negative and neutral emotional states may provide a clearer picture of the effects of self-rated emotion on visual working memory precision.

To address these concerns, we ran a preregistered (<https://osf.io/7jhgx>) high-powered replication ($N = 151$), having a sample of U.S. American participants recruited online via Prolific. To increase the power of our contrast, we included only the negative and neutral conditions. For this high-powered replication, we decided to use the same procedure as in our Experiment 3 for two reasons. First, we could use the IAPS images exactly as provided in the database (i.e., in color), which is probably better suitable to induce the desired emotional states. Second, the memoranda (orientation of triangles) is more emotionally neutral than colors which tend to be associated with certain emotional states.

Method

Participants

We performed an *a priori* power analysis using G*Power for a one-tailed *t*-test for paired samples. We assumed an effect of $d_z = 0.25$ (which is substantially smaller than the $d_z = 0.58$ and $d_z = 0.71$ reported by Xie and Zhang, 2016), and we aimed for power $1 - \beta = 90\%$ (with an error level of $\alpha = .05$). This resulted on a minimum sample size of $N = 139$. In our preregistration, we proposed to collect the data of $N = 150$ participants, and to increase this sample size to $N = 200$ in case the results were ambiguous (i.e., $BF < 3$) for or against our hypotheses.

Participants were recruited via Prolific, and we set the following recruitment criteria: (1) age between 18 and 35 years old; (2) English as first language, (3) normal vision; and (4) residency in the USA.

We obtained data of $N = 153$ participants (mean age = 26.5 years, 71 women, 81 men, and 1 other). Regarding education level, 32 participants reported having only a high-school degree, 101 reported having a college degree, 17 reported having a master's degree, and 3 reported having a doctor degree. We excluded data of one participant that never entered ratings for the images and one participant that reported being 40 years old, leaving a final sample of $N = 151$ for subsequent analyses. We assessed overall performance in the task (see Figure 13A), and the distribution of performance was bimodal. We identified 26 people that performed the task with relatively low accuracy (recall error $> 65^\circ$), and this value seemed to separate the two groups of individuals in our task. Because low performing individuals may cloud the contrast between conditions, we conducted our analysis once with the whole sample and once with this smaller sub-sample ($N = 125$). Results were similar for both sets of analyses.

Participants completed the task in about 40 min, and received as compensations 5.5 pounds. Before the start of the experiment, participants read an online informed consent form explaining the nature of the task, warning them that images with sensitive content would be presented thereby potentially eliciting emotional reactions, and explaining that the data would be posted online after anonymization. Participants clicked on statements in this form to indicate their consent to participation. At the end of the study, participants were debriefed regarding the purpose of the study.

An example of the experimental task (without the IAPS images which consist of protected content) is available in our preregistration page at <https://osf.io/7jhgx>.

Materials and Procedure

The experimental task and procedure were the same as described for Experiment 3, with the following exceptions. First, we programmed the task for online testing using lab.js (Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2020) which is an HTML and JavaScript based language. The experimental task was hosted in JATOS, and the link for data-collection was distributed via Prolific. Second, given that the study was run online, participants completed the experiment on their own laptop devices, and screen-size was unconstrained. To make the task suitable for somewhat smaller screens, the height of the triangles was set to 60 pixels. Third, the stimuli were presented across four fixed locations along an imaginary circle (radius = 120 pixels) set at 0°, 90°, 180°, and 270° degrees (i.e., cardinal locations), which was centered in the middle of the screen. Fourth, the IAPS images were presented across a rectangle measuring 553 x 400 pixels presented on the top-half of the screen, and the SAM scale appeared below it with dimensions of 670 x 106 pixels.

We selected the 60 negative images and the 60 neutral images used by Xie and Zhang (2016) for manipulating emotion, and we presented these images twice during the experiment. In total, participants completed 240 trials of the negative and neutral conditions which were randomly intermixed. Before the experiment proper, participants completed 6 practice trials (3 with negative and 3 with neutral images). The practice images were different from the ones used in the main experiment but were of the same average valence and arousal as compared to their respective categories.

The experimental task was divided into four blocks of 60 trials. After each block, a message was presented indicating the average recall error across all experimental trials completed so far. In the same screen, we also included a reminder message warning people to rate the images regarding how they made them feel at that moment. This block-message allowed participants to take a short break every 10 minutes.

Results

Figure 12

Mean Self-Rated Emotion in Experiment 6.

Note. Panel A shows data of the whole sample. Panel B shows the ratings for participants with an average recall error lower than 65°.

Self-Rated Emotion

Figure 12 presents self-rated emotion across valence conditions in Experiment 6 for the whole sample (Panel A) and for the sub-sample with recall error lower than 65° (Panel B). In both panels, self-rated emotion differed substantially between valence conditions (whole

sample, $BF_{10} = 9.63 \times 10^{65}$; sub-sample, $BF_{10} = 2.20 \times 10^{59}$).⁹ This indicates that our manipulation of valence was successful in modulating participant's self-rated emotional reactions.

Recall Error

Figure 13 depicts recall error across valence conditions when considering the whole sample (Panel B) and the sub-sample of participants with recall error $< 65^\circ$ (Panel C). For both sets, there was evidence that recall error did not differ as a function of valence condition (whole sample, $BF_{10} = 0.09$; sub-sample, $BF_{10} = 0.16$).

Figure 13

Recall Error in Experiment 6. Panel A Displays Individual Differences in Overall Recall Error.

Panels B and C Depict Recall Error as a Function of Valence Condition.

Note. For panels B and C, dots represent data of individual participants. The triangle shows the sample average in each condition. Error-bars depict 95% within-subjects confidence intervals.

⁹ The Supplementary Analyses available in the OSF shows that average ratings per image were closely related to the normative ratings used to select the IAPS images.

Mixture Modeling

As per the other experiments, we modeled the data of Experiment 6 using two approaches: the hierarchical mixture model implemented in the CatContModel package, and the MLE procedure implemented in Matlab (see Data Analysis Section in Experiment 1). We present both here, and note that there is general agreement between the two procedures with regards to the main conclusions.

Hierarchical Fitting. We fitted the model using 300'000 iterations, and we discarded the first 50'000 iterations as burn in. Figure 14 presents the group-level posterior means for the parameters. For the whole-sample, there was no evidence for differences between valence condition in memory imprecision, $BF_{10} = 0.091$, and in probability of recall, $BF_{10} = 0.090$. The same pattern holds when considering only data of the sub-sample with recall error $< 65^\circ$: memory imprecision, $BF_{10} = 0.086$, probability of recall, $BF_{10} = 0.099$.

MLE Fitting. Figure 15 presents the parameter estimates from the MLE model fitting which was applied to data of each individual in each condition separately. For memory imprecision, there was substantial evidence against lower imprecision in the negative condition compared to the neutral one when considering the whole sample, $BF_{10} = 0.14$ (one-tailed); and the sub-sample with recall error $< 65^\circ$, $BF_{10} = 0.07$ (one-tailed). For probability of recall, there was also evidence against differences between valence conditions (whole sample, $BF_{10} = 0.32$; sub-sample, $BF_{10} = 0.21$ – both two-tailed).

Figure 14

Group Level Parameter Estimates from the Hierarchical Mixture Model in Experiment 6.

Note. Error-bars depict the 95% credible interval of the parameter posterior.

Figure 15*MLE Parameter Estimates for Experiment 6.*

Note. Dots represent data of individual participants. The triangle shows the sample average in each condition. Error-bars depict 95% within-subjects confidence intervals.

Discussion

In Experiment 6, we conducted a high-powered replication of Experiment 3. We improved the methods in two regards: we collected data of residents in North America which may have a more similar background as the ones in Xie and Zhang (2016) – hence reducing cultural differences, and we only contrasted the negative and neutral valence conditions, reducing the overall length of the experiment, thereby mitigating fatigue and also the overall variation in affect. This study replicated the results we obtained in Experiments 1, 2, 3, and 5: there was no evidence that self-rated negative emotion changed either the quality or the quantity of the representations maintained in visual working memory in comparison to a neutral emotional state.

General Discussion

Can short-lived emotional states change how we select, process, and maintain neutral information in mind for ongoing cognitive activities? Recently, three studies indicated that self-rated negative emotional states (in comparison to either neutral or positive states) increased the precision with which information was stored within visual working memory (Long et al., 2020; Spachholz et al., 2014; Xie & Zhang, 2016). These findings were interpreted as indication that negative emotion changes processing style towards more detailed oriented processing or a narrower attentional focus (Easterbrook, 1959; Fredrickson & Branigan, 2005; Gable & Harmon-Jones, 2008, 2010; Uddenberg & Shim, 2015). Our repeated failures to replicate these effects across 6 studies conducted in four different countries (across three labs and one online platform) challenge the generalizability of this effect.

To manipulate emotion, we randomly varied whether negative, neutral, or positive images were presented before each trial of a visual working memory task that allowed the assessment of the quantity and quality of memory representations (Experiments 1-4, and 6). In Experiment 5, we combined the images with emotional music during an extended induction phase, and exposed participants to only one emotional state (either negative, neutral, or positive) prior to the visual working memory task.

Overall Effects Across the 7 Studies

Across our experiments, we collected data of a total of $N = 332$. Figure 16 presents an overall assessment of the effect of negative and positive self-rated emotion on memory imprecision (panel A) and probability of recall (panel B) across the seven experiments.

In six of the seven experiments we found no support for the hypothesis that self-rated negative emotion changed processing priorities within visual working memory towards more detailed processing thereby yielding more precise memories. We also found no credible effect of self-rated negative emotion on probability of recall. Only one condition in one experiment showed lower memory imprecision following both negative and positive self-rated emotion. Effects of positive emotion on recall probability were also a bit varied with most studies showing no change, and two studies showing a tendency for changes in opposite directions. Overall, these results challenge the assumption that brief changes in emotional state alter the quality and quantity of information stored in visual working memory.

Figure 16

Posterior of the Effect of Negative and Positive Self-Rated Emotion Across All Experiments.

Note. Dots represent the mean, and the line the 95% credible interval of the posterior. Red and blue lines indicate posterior differences that do not include 0.

Our findings add to an increasing literature showing mixed findings of emotional state on cognitive processing (for review Ribeiro et al., 2019). For example, there are mixed findings regarding the impact of self-rated negative emotion (e.g., sadness) upon attentional selection: some studies have found that self-rated sadness increases attentional focus (Basso, Schefft, Ris, & Dember, 1996; Fenske & Eastwood, 2003; Gasper & Clore, 2002; Melcher, Obst, Mann, Paulus, & Gruber, 2012), and some studies have found that sadness had no impact on attentional processing (Finucane, Whiteman, & Power, 2010; Fredrickson & Branigan, 2005; Rowe, Hirsh, & Anderson, 2007; Tibboel, 2018; Uddenberg & Shim, 2015; von Mühlenen, Bellaera, Singh, & Srinivasan, 2018) or even that it reduced it (Gable & Harmon-Jones, 2010). There is also conflicting evidence regarding the impact of positive emotion on attentional selection: some studies have reported that positive affect broadens attentional selection (Domachowska et al., 2016; Gable & Harmon-Jones, 2008), but other have failed to replicate these effects (Bruyneel et al., 2013; Grol & Raedt, 2014).

So far it is unclear which variables are responsible for these mixed findings. Accordingly, the field as a whole would benefit from large-scale replications (direct and conceptual ones) that aim at identifying critical conditions to observe effects of emotion on memory and attention. Moving forward, the sharing of materials and data will also be an important step in helping us determine the direction and robustness of emotion-cognition interactions.

Alternative explanations?

Our findings point to substantial evidence against an effect of mild-intensity self-rated emotional states on the processing of information in visual working memory. Are there

confounds that might explain our lack of effects? There are five potential criticisms or limitations that have been raised to our study.

First, one criticism that could be raised is that our participants might not have experienced the emotions they reported experiencing, and we used no other means to assess emotion but their self-report. We note that the reported valences in Experiment 1-4 and 6 were in clear agreement with the ones reported by Xie and Zhang (2016) and with the normative ratings from the IAPS. Furthermore, in Experiment 5 in which a more extended emotional induction phase was implemented, we only retained for analysis participants that reported a congruent emotion after the induction.

Self-reported measures are the most common means to assess the effectiveness of emotional induction (Joseph et al., 2020), and they were also used in the studies we attempted to replicate. It is worth noting that self-rated emotion has been considered a valid instrument to assess emotional experience when reports are collected in the moment the emotion is induced (in contrast to retrospective reports), and when emotion is rated in a dimension scale (valence or arousal) as used here (for a review see Mauss & Robinson, 2009). Accordingly, we see no reasons to doubt the validity of these self-reports just because we did not observe the predicted effects of emotional valence upon visual working memory. That said, we cannot exclude the possibility of demand characteristics (e.g., reporting a negative state when viewing a car crash image, even if participants feel nothing). In order to assess for this possibility, future studies could include manipulations of demand (Polivy & Doyle, 1980), i.e., manipulations of what the participants think the experimenter wants him to report. Furthermore, it may be important to take a multi-measure approach (e.g., eye-tracking, skin conductance, or event-

related potentials) to further validate the self-reported emotional state of the participants.

Autonomic responses and startle responses, for example, can be used to discriminate between valence and arousal dimensions (Mauss & Robinson, 2009). These manipulations can further shed light on the timeframe of the emotional response to the images, and complement the assessment of self-rated emotion.

Second, it is possible that self-rated negative emotion has an impact on the performance of some individuals but not others, and that our study simply did not sample individuals from the population that is sensitive to the variation in emotional state. One potential individual difference variable that could modulate the response to negative emotion is adverse childhood experiences (Anda et al., 2006) which can modulate responses to stress. Future studies should consider large scale replications and the inclusion of individual measures that are associated with the response to negative events.

Third, another criticism is that our study elicited broad classes of emotions categorized as negative, neutral, or positive in line with a circumplex model (Posner, Russell, & Peterson, 2005; Russell, 1980). The emotions elicited by the IAPS images were fairly heterogeneous within each valence category. For example, negative images were associated with sadness, disgust, fear. Positive images were associated with pleasure, fun, love, lust, cuteness, and so forth. As such, at the moment it is not clear if specific basic emotions (Ekman, 1992) could induce changes in visual working memory. Our findings can only speak to the replicability of the effects reported in the studies of Xie and Zhang (2016) and, to some extent, of Spachtholz et al. (2014). Investigating the impact of more homogeneous emotions remains a task for future studies.

Fourth, yet another criticism is that our studies used visual stimuli to manipulate emotion and assessed its effects upon visual working memory. In Experiment 5, we used music to induce emotion, but we still included images concurrently with the music. Hence so far it is unclear whether the effect of emotion on visual working memory could have been masked by interference created by presentation of the emotional images. Future studies may consider using non-visual forms of emotional induction.

Fifth, a final potential criticism relates to our choice of analysis method, i.e., to fit the mixture model of Zhang and Luck (2008) using an hierarchical Bayesian implementation, and whether this model is a fair representation of the data. We fitted the Zhang and Luck model because this model was the one used in the previously published studies on the topic and because this is one of the most influential analytical procedures for the type of visual working memory task employed here (Ma, 2018). We used a hierarchical Bayesian implementation because this type of implementation has several advantages over models fitted separately to the data of each participant, including improved recovery of model parameters in particular with low numbers of trials (Oberauer et al., 2017). We also fitted the data of all experiments using MLE, which allows for independent search of best fitting parameters for each individual in each condition – the results of this analysis generally agreed with the ones obtained with the hierarchical model. Notwithstanding, its widespread use, the Zhang and Luck mixture model has been criticized (Ma, 2018). We therefore also analyzed model-free performance scores (i.e., recall error), which provide a raw index of performance in the task that is not compromised by assumptions of how information ought to be represented in visual working memory. In none of the experiments, there was evidence that emotion affected performance in the task when using

any of these measures. Therefore, our conclusions are not bound by acceptance of the adequacy of this model as an adequate representation of the data.

What Does our Failure to Replicate Mean?

Our findings do not support the hypothesis that self-rated emotional states affect visual working memory. Using Bayesian inferential statistics, our findings do not simply point to a lack of a significant effect, but rather provide evidence for the absence of an effect. Indeed, we found strong evidence against an effect of valence across several experiments both in all three analyzed measures of visual working memory performance (raw recall error, imprecision, and recall probability).

Now, the question remains why did the published studies find an effect and we did not? Across the introductory and discussion sections of each experiment, we already reasoned that the evidence presented in the published studies was not particularly strong to start with. For example, Experiment 1 of Xie and Zhang (2016) only yielded a BF of around 2 in favor of the alternative hypothesis; their Experiment 2 provided a BF_{10} of 6. The study of Spachtholz et al. (2014) also suffered from limitations in how they assessed their data. These authors fitted the mixture model to the data of the entire group disregarding individual levels of performance. They argued that their relatively low number of trials ($N = 96$) was insufficient to yield robust estimates at the individual level. This is a highly unusual way to assess performance that leaves open whether the effects they reported are truly representative for their whole group of participants or rather reflect the impact of a few outliers. In the more recent study of Long et al. (2020), they suggested that the duration of array presentation (200 ms vs. 500 ms) could be critical to observe a trade-off in precision vs. capacity induced by self-rated negative emotion.

In line with those studies available at the time of running the present experiments, we have generally used the presentation duration of 400 ms (except for Experiment 5 in which this duration was of 1000 ms). Since this replicates the duration used by Xie and Zhang (2016), this should have provided sufficient time to produce an effect, but given that we haven't manipulated presentation duration over longer intervals, it is unclear whether with longer presentation duration, self-rated negative emotion could induce a trade-off on the allocation of visual working memory capacity.

Our failure to obtain an effect of self-rated emotion on visual working memory does not imply that emotion has no impact on cognition. Here, we asked participants to evaluate their emotional experiences along a continuum from negative to positive valence. This implies an understanding of these emotional experiences as two opposing polarities, in line with a circumplex model (Posner et al., 2005; Russell, 1980). This may provide an overly simplistic view of emotion (Cacioppo, Gardner, & Berntson, 1997). Future studies should consider the impact of emotion across different emotional conceptualizations. Note also that, again in line with the studies we aimed to replicate, our manipulation of emotion was very brief (in the order of a couple of seconds) and might be considered rather mild (presentation of an emotional image/emotional music). Stronger manipulations of emotion could have a reliably observable impact on visual working memory. Hence, future studies should also consider stronger emotional manipulations (e.g., of more personal value to the participant).

Given the relevance and ubiquity of emotional experiences, emotion-cognitive interactions deserve further investigation. Our experiments provide one step in the direction of assessing the impact of self-rated emotion on visual working memory quality and quantity.

Further studies are, however, needed to fully establish whether and when variations of self-rated emotional state affect processing priorities within visual working memory.

Conclusion: What Do Our Data Show About Emotion – Visual Working Memory Interactions?

Visual working memory keeps accessible the information needed for ongoing processing. To effectively fulfill this goal, visual working memory needs to be selective and to protect its contents from external and internal interference. Emotions are classically considered as information about the state of the individual, and from this perspective it would make sense to use it to change processing priorities within this system. That said, it would also make sense that visual working memory strives to avoid emotional biases, especially when the emotional state is irrelevant to the task goal, as it was the case here. Our observation that visual working memory was immune to moment-to-moment variations of emotion indicates that this system is actually fulfilling one of its core goals: to keep the information that guides our thoughts and actions protected from unduly external and internal influences (Liesefeld, Liesefeld, Sauseng, Jacob, & Müller, 2020).

These results resonate with the ones from a recent meta-analysis on the role of affective information in working memory which found that the overall effect of affective information on behavioral performance in working memory tasks has been between negligible and small (Schweizer et al., 2019). In line with the idea that working memory keeps information protected from affective biases, Schweizer et al. (2019)'s meta-analysis showed that affective information consistently activated brain regions related to processing of emotional information, yet producing small behavioral effects. This efficient control over the sources of biases, however, is substantially reduced in people with mental health problems; with this population

showing substantially larger effects of emotional information. Together these results indicate that in healthy participants, working memory will operate similarly under different emotionally-laden contexts, allowing us to keep unbiased track of task-relevant information.

References

- Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, Ch., Perry, B. D., ... Giles, W. H. (2006). The enduring effects of abuse and related adverse experiences in childhood. *European Archives of Psychiatry and Clinical Neuroscience*, 256, 174–186.
- Basso, M. R., Schefft, B. K., Ris, M. D., & Dember, W. N. (1996). Mood and global-local visual processing. *Journal of the International Neuropsychological Society*, 2, 249–255.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56, 893–897.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, 4, 561–571.
- Bekhtereva, V., & Müller, M. M. (2017). Bringing color to emotion: The influence of color on attentional bias to briefly presented emotional images. *Cognitive, Affective, & Behavioral Neuroscience*, 17, 1028–1047.
- Bergmann, H. C., Rijpkema, M., Fernández, G., & Kessels, R. P. C. (2012). The Effects of Valence and Arousal on Associative Working Memory and Long-Term Memory. *PLOS ONE*, 7, e52616.
- Bermudez, T., & Souza, A. S. (2017). Can emotional content reduce the age gap in visual working memory? Evidence from two tasks. *Cognition and Emotion*, 31, 1676–1683.

- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*.
<https://doi.org/10.1167/11.5.4>
- Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: The role of attention and motivation. *Emotion, 12*, 605–617.
- Bruyneel, L., van Steenbergen, H., Hommel, B., Band, G. P. H., De Raedt, R., & Koster, E. H. W. (2013). Happy but still focused: Failures to find evidence for a mood-induced widening of visual attention. *Psychological Research, 77*, 320–332.
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond Bipolar Conceptualizations and Measures: The Case of Attitudes and Evaluative Space. *Personality and Social Psychology Review, 1*, 3–25.
- Constant, M., & Liesefeld, H. R. (2020). Massive effects of saliency on information processing in visual working memory. *Psychological Science*. Retrieved from
<https://psyarxiv.com/q36zu/>
- Curby, K. M., Smith, S. D., Moerel, D., & Dyson, A. (2019). The cost of facing fear: Visual working memory is impaired for faces expressing fear. *British Journal of Psychology, 110*, 428–448.
- Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science, 298*, 1191–1194.
- Domachowska, I., Heitmann, C., Deutsch, R., Goschke, T., Scherbaum, S., & Bolte, A. (2016). Approach-motivated positive affect reduces breadth of attention: Registered replication

- report of Gable and Harmon-Jones (2008). *Journal of Experimental Social Psychology*, 67, 50–56.
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66, 183–201. (1961-03074-001).
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fenske, M. J., & Eastwood, J. D. (2003). Modulation of Focused Attention by Faces Expressing Emotion: Evidence From Flanker Tasks. *Emotion*, 3, 327–343.
- Finucane, A. M., Whiteman, M. C., & Power, M. J. (2010). The Effect of Happiness and Sadness on Alerting, Orienting, and Executive Attention. *Journal of Attention Disorders*, 13, 629–639.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10. <https://doi.org/10.1167/10.12.27>
- Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, 19, 313–332.
- Gable, P. A., & Harmon-Jones, E. (2008). Approach-Motivated Positive Affect Reduces Breadth of Attention. *Psychological Science*, 19, 476–482.
- Gable, P. A., & Harmon-Jones, E. (2010). The Blues Broaden, but the Nasty Narrows: Attentional Consequences of Negative Affects Low and High in Motivational Intensity. *Psychological Science*, 21, 211–215.

- Galinha, I. C., & Pais-Ribeiro, L. (2005). Contribuição para o estudo da versão portuguesa da Positive and Negative Affect Schedule (PANAS): I - Abordagem teórica ao conceito de afecto. *Análise Psicológica*, 23, 209–218.
- Galinha, I. C., Pereira, C. R., & Esteves, F. (2014). Versão reduzida da escala portuguesa de afeto positivo e negativo - PANAS-VRP: Análise fatorial confirmatória e invariância temporal. *Psicologia*, 28, 50–62.
- Gasper, K., & Clore, G. L. (2002). Attending to the Big Picture: Mood and Global Versus Local Processing of Visual Information. *Psychological Science*, 13, 34–40.
- Gray, J. R. (2004). Integration of Emotion and Cognitive Control. *Current Directions in Psychological Science*, 13, 46–48. Scopus.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, 15, 1176–1194.
- Grol, M., & Raedt, R. D. (2014). Effects of positive mood on attentional breadth for emotional stimuli. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01277>
- Hardman, K. O. (2016). CatContModel: Categorical and Continuous Working Memory Models for Delayed Estimation Tasks (Version Version 0.7.1). Retrieved from <https://github.com/hardmanko/CatContModel/releases/tag/v0.6.1>
- Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43, 30–54.

- Haupt, M., Ruiz-Rizzo, A. L., Sorg, C., & Finke, K. (2019). Phasic alerting effects on visual processing speed are associated with intrinsic functional connectivity in the cingulo-opercular network. *NeuroImage*, 196, 216–226.
- Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2020). *lab.js: A free, open, online experiment builder*. Zenodo.
- Jackson, M. C., Linden, D. E. J., & Raymond, J. E. (2014). Angry expressions strengthen the encoding and maintenance of face identity representations in visual working memory. *Cognition & Emotion*, 28, 278–297.
- Jackson, M. C., Wolf, C., Johnston, S. J., Raymond, J. E., & Linden, D. E. J. (2008). Neural Correlates of Enhanced Visual Short-Term Memory for Angry Faces: An fMRI Study. *PLOS ONE*, 3, e3536.
- Jackson, M. C., Wu, C.-Y., Linden, D. E., & Raymond, J. E. (2009). Enhanced visual short-term memory for angry faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 363–374.
- Jeffreys, H. (1961). *Theory of probability*. Oxford: UK Oxford University Press.
- Jingling, L., & Zhaoping, L. (2008). Change Detection is Easier at Texture Border Bars When They are Parallel to the Border: Evidence for V1 Mechanisms of Bottom-up Salience. *Perception*, 37, 197–206.
- Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, 146, 355–375. (2020-04359-001).

- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773–795.
- Kensinger, E. A. (2007). Negative Emotions Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence. *Current Directions in Psychological Science*, 16, 213–218.
- Kensinger, E. A., & Corkin, S. (2003). Effect of Negative Emotional Content on Working Memory and Long-Term Memory. *Emotion*, 3, 378–393.
- Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43, 149–164.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. Technical Report*. Gainesville, FL.
- Laux, L., Hock, M., Bergner-Köther, R., Hodapp, V., & Renner, K.-H. (2013). *Das State-Trait-Angst-Depressions-Inventar—Manual*. Göttingen: Hogrefe.
- Liesefeld, H. R., Liesefeld, A. M., Sauseng, P., Jacob, S. N., & Müller, H. J. (2020). How visual working memory handles distraction: Cognitive mechanisms and electrophysiological correlates. *Visual Cognition*, 0, 1–16.
- Long, F., Ye, C., Li, Z., Tian, Y., & Liu, Q. (2020). Negative emotional state modulates visual working memory in the late consolidation phase. *Cognition and Emotion*, 0, 1–18.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17, 391–400.
- Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model. *BioRxiv*, 268961.

- Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 369–380.
- Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, 6, 114–133.
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, 23, 209–237.
- Melcher, T., Obst, K., Mann, A., Paulus, C., & Gruber, O. (2012). Antagonistic modulatory influences of negative affect on cognitive control: Reduced and enhanced interference resolution capability after the induction of fear and sadness. *Acta Psychologica*, 139, 507–514.
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.12-2). Retrieved from <http://CRAN.R-project.org/package=BayesFactor>
- Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science*, 24, 550–556.
- Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, 51, 45–100.
- Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H.-Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, 17, 11–11.

- Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion–cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00058>
- Pertsov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1224–1231.
- Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, 13, 160–166.
- Polivy, J., & Doyle, C. (1980). Laboratory induction of mood states through the reading of self-referent mood statements: Affective changes or demand characteristics? *Journal of Abnormal Psychology*, 89, 286–290. (1980-11238-001).
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17, 715–734.
- Quintão, S., Delgado, A. R., & Prieto, G. (2013). Validity study of the Beck Anxiety Inventory (Portuguese version) by the Rasch Rating Scale model. *Psicologia: Reflexão e Crítica*, 26, 305–310.
- R core team. (2017). *R: A language and environment for statistical computing*. Vienna: Austria: R Foundation for Statistical Computing.

- Ribeiro, F. S., Santos, F. H. dos, & Albuquerque, P. B. (2019). How does allocation of emotional stimuli impact working memory tasks? An overview. *Advances in Cognitive Psychology*, 15, 155–168.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21, 301–308.
- Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, 25, 102–113.
- Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences*, 104, 383–388.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25, 128–142.
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22, 322–339. (2015-56330-001).
- Schwarz, N., & Clore, G. L. (2003). Mood as Information: 20 Years Later. *Psychological Inquiry*, 14, 296–303.

- Schweizer, S., Satpute, A. B., Atzil, S., Field, A. P., Hitchcock, C., Black, M., ... Dalgleish, T. (2019). The impact of affective information on working memory: A pair of meta-analytic reviews of behavioral and neuroimaging evidence. *Psychological Bulletin*, 145, 566–609. (2019-22428-001).
- Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., & Pureza, R. (2015). Adaptation of the International Affective Picture System (IAPS) for European Portuguese. *Behavior Research Methods*, 47, 1159–1177.
- Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 78, 1839–1860.
- Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention, Perception, & Psychophysics*, 79, 1897–1916.
- Souza, A. S., Rerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics*, 76, 2080–2102.
- Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1237–1256.
- 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.

- 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.
- Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, 166, 277–297.
- Spachtholz, P., Kuhbandner, C., & Pekrun, R. (2014). Negative affect improves the quality of memories: Trading capacity for precision in sensory and working memory. *Journal of Experimental Psychology: General*, 143, 1450–1456.
- Tabi, Y. A., Husain, M., & Manohar, S. G. (2019). Recall cues interfere with retrieval from visuospatial working memory. *British Journal of Psychology*, 110, 288–305.
- Thomas, P., Jackson, M. C., & Raymond, J. E. (2014). A threatening face in the crowd: Effects of emotional singletons on visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 253–263.
- Tibboel, H. (2018). Mood effects on attentional control: A preregistered replication study and critical analysis. *Cognition and Emotion*, 32, 145–157.
- Uddenberg, S., & Shim, W. M. (2015). Seeing the world through target-tinted glasses: Positive mood broadens perceptual tuning. *Emotion*, 15, 319–328.
- Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, 123, 394–409. (1995-08699-001).
- Västfjäll, D. (2001). Emotion induction through music: A review of the musical mood induction procedure. *Musicae Scientiae*, 5, 173–211.

- Vaz Serra, A. S., & Pio Abreu, J. L. (1973). Aferição dos quadros clínicos depressivos. II-Estudo preliminar de novos agrupamentos sintomatológicos para complemento do Inventário Depressivo de Beck. *Coimbra Médica*, 20, 713–736.
- von Mühlenen, A., Bellaera, L., Singh, A., & Srinivasan, N. (2018). The effect of sadness on global-local processing. *Attention, Perception, & Psychophysics*, 80, 1072–1082.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070. (1988-31508-001).
- Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19, 1057–1064.
- Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin & Review*, 20, 348–355.
- Xie, W., & Zhang, W. (2016). Negative emotion boosts quality of visual working memory representation. *Emotion*, 16, 760–774.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235.
- Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, 22, 1434–1441.