Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 2, 321–347
https://doi.org/10.1037/xge0001084

Verbal Descriptions Improve Visual Working Memory but Have Limited Impact on Visual Long-Term Memory

Clara Overkott[1] and Alessandra S. Souza[1, 2]

[1] Department of Psychology, University of Zurich
[2] Faculty of Psychology and Educational Sciences, University of Porto

How do verbal descriptions affect visual memory over the short and long term? Here we show for the first time that verbal labeling can boost visual memories, but the source of this benefit depends on whether representations are maintained over the short term in visual working memory or over the long term in visual long-term memory. Across three experiments, we contrasted color memory of randomly colored objects when participants labeled (a) the color, (b) the object, or (c) the color-object binding, to memory under an articulatory suppression condition inhibiting labeling. Memory was tested at two time points: after three objects (visual working memory) and at the end of the experiment (visual long-term memory). In Experiment 1, color labeling improved, whereas object labeling impaired, visual working memory in comparison to suppression. Visual long-term memory remained unchanged across conditions. Experiment 2 tested whether this was attributable to poor overall long-term learning by repeating the colored objects over three successive working memory trials. This increased performance over the short and long term, yet labeling did not change learning rate over repetitions or delayed memory performance, showing no long-term memory benefit. In Experiment 3, a labeling benefit was observed when the color-object binding was labeled both over the short and long term. Mixture modeling indicated that color-labeling benefits in visual working memory resulted from an increase of detailed visual memory, whereas long-term memory benefits accrued from categorical representations. Our findings point to dissociations on the role of language in visual working memory and visual long-term memory.

Keywords: labeling, language, long-term memory, mixture modeling, visual working memory

How do verbal descriptions affect visual memory over the short and long term? We may describe the visual information needed for ongoing processing (e.g., the positions of the cars approaching us while changing lanes), or information to be used over longer periods (the route we will take to arrive at a certain place). Retention of visual information over short and long time scales are supported by different memory systems. Visual working memory is the system that keeps visual information available for ongoing cognition. Visual working memory has a limited capacity, and therefore people can only maintain a small amount of information in this system at a given time (Luck & Vogel, 2013; Oberauer et al., 2016). In contrast, visual long-term memory stores large amounts of visual information over long periods of time, varying from several minutes to years, with no upper limit on how much information can be committed to it (Brady et al., 2008; Konkle et al., 2010a, 2010b).

Verbal labeling has been found to improve visual working memory (Forsberg et al., 2020; Souza & Skóra, 2017) by increasing the fidelity of the representations stored in this system. In contrast, labeling has been reported to be inconsequential for visual long-term memory (Kelly & Heit, 2017): labeling produced neither a benefit nor a cost to memory performance over the long term. What are the reasons for these discrepant findings? The present study aimed to provide a first systematic comparison of how labeling affects visual representations retained in visual working memory for an immediate task goal and retained in visual long-term memory for delayed recall.

In the following, we first review how memories are retained over the short and long term in relation to the quantity and quality of the information stored. Next, we describe how labeling has been linked to categorical knowledge and current hypotheses on how labeling changes visual representations. Finally, we discuss whether there are reasons to suspect that labeling operates

This article was published Online First September 13, 2021.
Clara Overkott https://orcid.org/0000-0002-8954-7381
Alessandra S. Souza https://orcid.org/0000-0002-1057-8426
This research was support by a Grant from the Swiss National Science Foundation (100019_169302) awarded to Alessandra S. Souza. We thank Gian-Luca Gubler and Vanessa Vallesi for helping with data collection. All experiments reported in this article were preregistered in the Open Science Framework. Preregistrations, materials, data, and analysis scripts can be found at https://osf.io/pn3fh/.
Correspondence concerning this article should be addressed to Clara Overkott, Department of Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zürich, Switzerland. Email: claraoverkott@gmail.com

differently when memories are stored in visual working memory versus visual long-term memory and then delineate our research aims.

Visual Memories Over the Short and Long Term

Memories stored in visual working memory and visual long-term memory differ in several regards. Research over the past 10 years has demonstrated that visual memories can be described in terms of parameters reflecting its quantity and quality by using mixture models (Zhang & Luck, 2008). In these models, quantity refers to the number of objects accessible for recall, whereas quality refers to the fidelity or precision with which these objects are stored. This approach is commonly applied in the so-called fidelity tasks where participants are required to reproduce, using a continuous scale, one of the features of the memoranda (Prinzmetal et al., 1998; Wilken & Ma, 2004; Zhang & Luck, 2008). For example, a participant may be instructed to remember the precise color of a set of real-world objects. At test, the object is presented in gray, and the task is to reproduce the color associated with the object using a continuous color wheel. This task has been used to examine changes in the accessibility and precision of features of a small set of objects maintained in visual working memory in comparison to the features of hundreds of objects stored in long-term memory (Biderman et al., 2019; Brady et al., 2013). Biderman et al. (2019) showed that both memory precision and the probability of memory retrieval were higher when information was maintained in visual working memory than in long-term memory (see also Miner et al., 2020). This shows that visual working memory maintenance confers higher accessibility and fidelity to visual representations.

More recently, these mixture models have been extended to incorporate parameters reflecting the contribution of categorical knowledge to memory (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017; Persaud & Hemmer, 2016). This is because systematic categorical bias has been uncovered when features are reproduced from perception (Bae et al., 2015), visual working memory (Donkin et al., 2015; Hardman et al., 2017), and visual long-term memory (Persaud & Hemmer, 2016). In a nutshell, a substantial proportion of responses in fidelity tasks are influenced by the category the memorized feature belongs to (e.g., "red") rather than the specific feature-value studied (e.g., the specific reddish hue).

Here we will use a categorical-continuous mixture model (Hardman et al., 2017) to probe how conditions prompting and preventing verbal labeling change parameters associated with the storage of categorical and continuous information in visual working memory and long-term memory. Implicitly, categorical knowledge has been related to verbal labeling, whereas continuous information was associated with purely visual memory limitations. In the next section, we present the available evidence for the labeling effects on visual memory over the short and long term and how labeling affects categorical and continuous memory parameters.

Labeling Versus Categorical Representations

Although categorical representations are usually assumed to reflect the impact of verbal labeling on visual memory, this assumption has empirically been underinvestigated. Recently, Souza and Skóra (2017) manipulated labeling opportunities in a visual working memory fidelity task: participants studied four sequentially presented colored dots while either (a) labeling the presented colors aloud or (b) saying "bababa" aloud (a verbal suppression procedure that inhibits labeling). During test, the colors of all four dots were reproduced on a color wheel. The authors observed that color labeling improved recall performance in comparison to suppression. Mixture modeling revealed that color labeling increased the tendency to respond categorically as opposed to guessing. This is in line with the assumption that verbal labels provide categorical information. Surprisingly, labeling also impacted continuous memory by either increasing the proportion of continuous memory responses as opposed to guessing or its precision. This effect was interpreted as an indication that labeling activated categorical information in visual long-term memory, thereby augmenting or protecting continuous representations held in visual working memory. Forsberg et al. (2020) replicated this study comparing performance between younger and older adults. They also observed that color labeling improved categorical and continuous visual working memory among the younger adults. However, older adults showed a benefit only in the storage of categorical information in visual working memory. The protection afforded by labeling for the storage of continuous details, therefore, may be subjected to age-related cognitive decline.

One may wonder whether the effect of labeling described above is related to a benefit of using verbal labels or a cost induced by the articulatory suppression procedure. There is no evidence that articulatory suppression impacts visual working memory. In a comprehensive test, Sense et al. (2017) did not observe any cost of articulatory suppression compared with silent study in a change-detection task requiring recognition of a change between two visual arrays. Given that change-detection tasks use brief study duration combined with short retention intervals, they provide little opportunity for verbalization of the memoranda. Hence any cost of suppression unrelated to labeling should be apparent in this task, but this was not the case. Souza and Skóra (2017) also provided a test of this possibility in their Experiment 4. They presented arrays of colored dots simultaneously for a brief interval, and tested memory after a 1-s or 3-s retention interval. This task was completed under articulatory suppression, silence, or overt color labeling. Performance in the suppression condition remained unchanged with increases in the retention interval. The silence and overt color labeling conditions, in contrast, did not differ from suppression in the 1-s retention condition but showed substantially better performance in the 3-s retention condition. Souza and Skóra argued that a short retention interval combined with presentation of multiple items hinders labeling, and hence imposing suppression or not makes little difference. Only when participants have sufficient time to label the items (overtly or covertly), performance improves. These findings support the conclusion that labeling produces a benefit.

In contrast to the labeling benefit in visual working memory, Kelly and Heit (2017) found that labeling was not unique in improving recognition performance in a visual long-term memory test. In their experiments, participants were presented with a series of colored objects (red or green) and were asked (a) to categorize the colors of objects as being either red or green, or to judge whether (b) they liked the presented object (preference judgment) or (c) the object was living/nonliving (animacy judgment). The specific hue of red or green was irrelevant for the categorization decisions. Visual long-term memory for the specific object color-hue was then assessed in a surprise test at the end. Categorizing the object in regard to its color resulted in a shift toward fewer categorical color responses in the memory test than when participants made preference or animacy judgements. Critically, this did not increase the probability of choosing the correct color. This decrease in categorical responses was also found when foreknowledge of the upcoming visual long-term memory test was given in all conditions. The authors concluded that color labeling reduced categorical bias, but this facilitation was not unique to labeling.

To summarize, these two studies suggest contrasting effects of color labeling on the retention of color in visual working memory and visual long-term memory. Labeling benefited visual working memory by increasing access to both continuous and categorical information (Forsberg et al., 2020; Souza & Skóra, 2017). In contrast, labeling reduced categorical bias in a visual long-term memory test, but this did not increase memory for the correct color (Kelly & Heit, 2017). These divergent findings may suggest that visual working memory and visual long-term memory are affected differently by verbal labeling. The caveat here is that these two studies manipulated verbal labeling differently. In the former, participants were explicitly instructed to say the colors aloud, whereas in the latter participants were not instructed to overtly label the colors (they categorized them via keypress). Hence, it is unclear whether participants were relying on verbal labels to perform the categorization task after a few trials. These divergent findings may therefore reflect differences in the procedure assumed to generate labeling behavior. Another critical difference across these studies refers to the memory test. In the study of Souza and Skóra (2017), participants reproduced the colors using a continuous color wheel. In the study of Kelly and Heit (2017), participants reported the remembered colors by picking it from a five-choice alternative set. The latter procedure is limited in the assessment of memory precision and might therefore reduce the chance of measuring a labeling benefit. Accordingly, before we can conclude that labeling affects visual working memory and visual long-term memory differently, these two systems need to be compared under equivalent conditions. This will be one of the main goals of the present study.

Before introducing to the empirical work, it is important to understand the proposed mechanisms by which labeling can influence visual memories. Several hypotheses have been raised, which are reviewed in the following section.

Hypotheses of the Labeling Effect

Here, five hypotheses will be discussed that make different predictions regarding how labeling affects storage of categorical and continuous representations. Essentially, none of these hypotheses make differential predictions regarding the role of language in visual working memory versus visual long-term memory, and most of them have received support from research evaluating either of these memory systems. This is probably the case because the effects of labeling on visual working memory and long-term memory have not been put in direct comparison before.

Hypothesis 1: Verbal Recoding

The *verbal recoding hypothesis* (Souza & Skóra, 2017) or *label distorting memory hypothesis* (Kelly & Heit, 2017) assumes that during encoding verbal labeling creates a verbal trace at the expense of the visual information. For example, labeling the picture of a light-blue shoe as "blue" creates a verbal trace of "blue" whereas the visual details about the specific hue (e.g., shade of light blue) are lost. This hypothesis therefore predicts a cost of labeling for detailed visual memory.

Evidence for the verbal recoding hypothesis stems from the *verbal overshadowing* effect in visual long-term memory (Schooler & Engstler-Schooler, 1990). In the classical studies by Schooler and Engstler-Schooler, describing a face or a color interfered with recognition of the stimulus in a visual long-term memory test (see also Alogna et al., 2014; Brandimonte et al., 1997).

Memory distortion caused by labeling was also found by Lupyan (2008). In his study, participants were asked to label objects as belonging to either one of two categories (e.g., chair vs. lamp) or to rate their preference for one of the objects. Long-term recognition performance was impaired in the labeling condition in comparison to preference rating. Lupyan interpreted these findings as indicating that labeling caused the visual representations to drift toward the category prototype (see also Carmichael et al., 1932).

Hypothesis 2: Dual Trace

The *dual-trace hypothesis* (Souza & Skóra, 2017) assumes that labeling builds two memory traces: a verbal trace based on the verbal label that was assigned to the object and a visual trace of the object itself. This hypothesis stands in contrast to the verbal recording hypothesis, where labeling is assumed to generate only one verbal (categorical) trace. This hypothesis predicts that labels help memory by providing an additional source of categorical information, without changing the retention of the visual trace. Evidence for this hypothesis stems from Paivio's (1971, 1990) *dual coding* model: Visual information has an advantage because it can be encoded in two formats, namely as a visual representation and as a verbal label.

This assumption is exemplified in the modeling implemented by Donkin et al. (2015): They included verbal labeling as a further component into a mixture model estimating the quantity and quality of visual working memory representations. Their modeling showed that the inclusion of this parameter better predicted their visual working memory data, because some responses seemed to have been guided by information provided by the label. Their modeling, however, does not assume that labeling induces any change in the visual trace.

Further evidence for the dual-trace hypothesis was found in visual long-term memory studies showing that the verbal overshadowing effect could be modulated or even reversed (Brandimonte et al., 1997; Brown et al., 2014). For example, Brown et al. (2014) asked participants to learn easy-to-label and hard-to-label pictures, with the assumption that participants would covertly label the easy-to-label pictures. Then, participants were asked to either provide a detailed description of the learned feature or do a filler task. The final memory test was meant to either favor retrieval of featural or global information of the object. These authors found that covert verbal labeling of the easy-to-label pictures impaired visual long-term memory performance, as would be predicted by the verbal overshadowing effect (see also, Brandimonte et al., 1992). However, a detailed description of the feature benefited visual long-term memory performance in a featural memory test. This provides evidence that the verbal overshadowing effect for labels can be reversed with feature descriptions that match the final memory test. This finding challenges the verbal recording hypothesis by showing that participants may have both the visual and the verbal traces accessible.

Hypothesis 3: Distinctiveness

The third hypothesis proposes that verbal labels make memory representations more distinct (Blanco & Gureckis, 2013; Kelly & Heit, 2017; Richler et al., 2013; Souza & Skóra, 2017). This *distinctiveness hypothesis* assumes that a label serves as an additional retrieval cue to the memory object or as a cue to augment encoding specificity (Blanco & Gureckis, 2013; Richler et al., 2011; Tulving & Thomson, 1973), thereby facilitating memory retrieval. Critically, if labels simply provide a distinctive cue to memory, it should not matter what type of label is used, as long as it provides a unique means to access the visual trace.

In verbal studies, a distinctiveness effect has been observed when comparing memory for words read aloud versus silently during study (MacLeod, 2010; MacLeod et al., 2010; Ozubko & Macleod, 2010). For visual long-term memory, Richler et al. provided some evidence for a distinctiveness benefit: They presented exemplars from either unique categories or exemplars sampled from only two categories. They showed that vocally labeling the unique categories during study yielded similar memory performance as a preference rating task. In contrast, the two-category labels impaired memory performance. Additionally, preference ratings using a 5-point scale during encoding provided more distinctiveness and presumably deeper processing than the labeling of the memory items with two categories (Blanco & Gureckis, 2013). These studies suggest that the uniqueness of a category label is essential for a distinctiveness benefit: The more unique, the better. Souza and Skóra (2017) also tested whether distinct labels could improve visual working memory for colors. They instructed participants to label the presentation order of a sequence of four colors (e.g., first, second, third, and fourth) under the assumption that these labels would increase distinctiveness in comparison to a condition with articulatory suppression. However, labeling their serial position did not provide any advantage.

Hypothesis 4: Activation of Categorical Visual Long-Term Memory

The *activation of categorical visual long-term memory hypothesis* (Souza & Skóra, 2017), based on the *label-feedback hypothesis* (Lupyan, 2012), assumes that verbal labels activate categorical knowledge in visual long-term memory. In this case, two visual traces are produced: one from visually encoding the object and the other is the visual long-term memory representation of the category activated by the verbal label. Activation of the visual categorical representation may allow data compression (see also, Brady et al., 2009): Instead of storing all of the details regarding the visual object, the memory trace may represent deviations in relation to the category, thereby reducing memory load. Accordingly, this hypothesis predicts a labeling benefit with more visual details being stored in memory. Evidence for a labeling benefit of this sort has been obtained by Souza and Skóra (2017): They showed that verbally labeling colors improved visual working memory compared with a suppression condition because of increases in categorical and continuous memory.

Further support for this hypothesis stems from studies finding that labels more efficiently cued the category (e.g., dog) of an object than nonverbal stimuli (e.g., a barking sound), thereby facilitating categorization and perceptual decisions (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015; Forder & Lupyan, 2019; Lupyan & Thompson-Schill, 2012).

Hypothesis 5: Cue to Focus Attention

Labels can also be viewed as a *cue to focus attention* in certain aspects of the visual object (Kelly & Heit, 2017). This hypothesis predicts that labeling may only be useful if it guides attention to relevant features, whereas it may be costly if it guides attention to irrelevant features. Critically, Kelly and Heit proposed that if attention is guided to the labeled feature irrespective of labeling, then labeling should be inconsequential. Kelly and Heit (2017) found that color labeling during study reduced color bias toward the color prototype in a surprise visual long-term memory recognition test in comparison to conditions that required an animacy judgment or preference rating during study. They argued that this occurred because the label guided attention to the relevant feature during study for the later memory test. When participants were informed about the relevant feature for the test before study, the advantage of color labeling vanished.

The Present Study

The main goal of the present study was to examine the impact of verbal labeling on both visual working memory and visual long-term memory using a color fidelity task. To the best of our knowledge, no previous study considered the impact of labeling concurrently on these two memory systems. Although the hypotheses of the labeling effect do not make differential predictions for retention over short and long timescales, there is empirical reason to suspect that labeling affects visual working memory and visual long-term memory differently. For example, whereas Souza and Skóra (2017) found a benefit of color labeling to retention of visual details in visual working memory, Kelly and Heit (2017) found neither benefits nor costs of color labeling in a visual long-term memory test. These findings are difficult to directly compare, however, because their experimental set-up differed in many regards. Accordingly, it is not clear to what degree their contradicting results reflects aspects of the experimental procedure versus true differences on the creation of visual memory representations to be used for ongoing cognition (e.g., in visual working memory) versus for later recall (e.g., in visual long-term memory). Here we designed a task to measure both memory systems using the same type of overt labeling manipulation and task requirements. This allowed us to directly examine how verbal labels influence the creation of memory representations to be accessed over the short and long term and to test predictions of the labeling hypotheses delineated above.

Given that the labeling hypotheses do not differentiate between visual working memory and visual long-term memory storage, this leads to the expectation that whatever mechanism operates over the short term should also affect performance over the long term. Our experiments provide a unique opportunity to address whether this is indeed the case. If the effect of labeling differs between visual working memory and visual long-term memory, this would require a revision of the labeling hypotheses and would support the separation of these two memory systems as independent of each other (Brady et al., 2011).

The general procedure of our experiments was as follows. We implemented two phases: a visual working memory phase containing the labeling manipulations, followed by a final delayed memory test that comprised our visual long-term memory phase. In the visual working memory phase, participants completed several trials of a continuous color fidelity task. Trials consisted of the sequential presentation of three colored objects. To assess the effect of verbal labeling on memory, participants were instructed to either (a) label the color (Experiments 1 and 2), (b) label the object (Experiment 1), or (c) label the color-object combination (Experiment 3). As a control condition in all experiments, participants also performed the task while saying "bababa" aloud (suppression) thereby inhibiting the use of verbal labeling.

At the visual working memory test, participants were tested on their memory for the colors of all three objects: they were shown the object in gray as a retrieval cue, and they were asked to reproduce its color by using a color wheel. After the end of the visual working memory phase, participants were asked to reproduce the color of all objects studied again (visual long-term memory phase). Our goal was to examine whether retrieval of an object's color in the visual long-term memory test would vary depending on the labeling manipulations implemented during the visual working memory phase. This allowed us to test whether labeling would affect memory representations similarly when they were retrieved from visual working memory and from visual long-term memory.

To foreshadow our results, we found a benefit of labeling the color and a cost of labeling the objects for the retention of color-object combinations in visual working memory in Experiment 1. There was no effect of labeling on visual long-term memory, independently of whether participants had foreknowledge about the visual long-term memory test (Experiment 1b) or not (Experiment 1a). However, overall performance in the visual long-term memory test was quite poor. To improve visual long-term memory learning, in Experiment 2, each trial of the visual working memory phase was repeated three times to increase long-term learning. Additionally, participants were only required to label the colors or to perform suppression (the object labeling condition was dropped). Across the three repetitions, performance improved in the visual working memory test thereby showing a learning effect. There was a color labeling benefit in visual working memory for the very first presentation of the color-object binding, but this benefit vanished over the course of the repetitions. Although performance improved overall in the final test, Experiment 2 showed no labeling effect in visual long-term memory replicating Experiment 1. In Experiment 3, participants were asked to label the color-object combinations (instead of only the color or only the object) and this was contrasted to suppression. For the first time across our series of experiments, we showed a labeling benefit in both visual working memory and in visual long-term memory.

Overall, we found evidence for a dissociation of the labeling benefit between the short term and the long-term. Modeling further showed that labeling benefited continuous memory over the short term, whereas this benefit was categorical in the long term. This indicates that the labeling benefit has different sources in visual working memory and visual long-term memory.

Experiment 1

The goal of Experiment 1 was to investigate whether the beneficial effect of color labeling in visual working memory would translate into better color memory in visual long-term memory. In addition, we included an object labeling condition that allowed us to further distinguish between the predictions of the labeling effect.

In the present experiment, participants were asked to (a) say "bababa" aloud thereby inhibiting labeling, or (b) label the color or (c) the shape of visual objects during the visual working memory phase. At the end of the study, they were then tested again on the same visual objects in a delayed memory test (visual long-term memory phase). The memory test in the visual working memory and visual long-term memory phases required participants to reproduce colors using a continuous color wheel. The use of a continuous color test allowed us to assess how labeling affected the storage of continuous and categorical information in both memory systems using a mixture modeling approach.

The five hypotheses of the labeling effect make differential predictions for the data of Experiment 1, which are summarized in Table 1. The *label recording* hypothesis (Hypothesis 1) predicts a labeling cost compared with the suppression baseline. This cost should be reflected on memory precision in the color labeling condition as the label replaces the fine-grained detail of the color hue. In the object labeling condition, in contrast, it should be reflected on the accessibility of the memory representation because the object's name would overshadow the color information. The *dual-trace* hypothesis (Hypothesis 2) predicts an increase in categorical responding as a function of color labeling with no change in continuous information. Object labeling should have no effect on memory performance, because this label lacks in providing information to improve color recall; The *distinctiveness* hypothesis (Hypothesis 3) predicts that labeling should increase the chance of recalling the visual information, and this increase should be larger for object than color labeling given that object labels provide a

Table 1
Summary of Predictions of the Labeling Hypotheses to the Data of Experiment 1

Hypothesis	Color labeling	Object labeling
1. Label recoding	↓ Memory precision	↓ Memory accessibility
2. Dual trace	↑ Categorical responses	=
3. Distinctiveness	↑ Accessibility	↑↑ Accessibility
4. Activation of categorical visual long-term memory	↑ Continuous memory	=
5. Cue to focus attention	=	↓ Memory

more unique cue to the memory representation. The *activation of categorical visual long-term memory* hypothesis (Hypothesis 4) predicts that labeling yields a benefit to categorical and continuous visual information. This benefit should only be observed to the color labels, because they are the only ones that activate the relevant categories to the memory test. Lastly, the *cue to focus attention* (Hypothesis 5) predicts that color labeling should be inconsequential as participants were already fully aware that color information was the relevant feature. Object labeling, in contrast, should lead to a cost because it draws attention away from the relevant feature for the test.

We ran two experimental versions. In Experiment 1a, participants were not informed about the visual long-term memory phase, whereas in Experiment 1b, participants were informed about the visual long-term memory phase at the beginning of the study. Our reasoning to disclose the occurrence of the visual long-term memory test in Experiment 1b was to motivate participants to try to remember the objects over the long-term, thereby possibly increasing visual long-term memory performance.

The research questions, method, and statistical hypotheses for Experiment 1a were preregistered and can be found at https://osf.io/wru4z/. Note that our preregistration was only concerned with differences between visual working memory and visual long-term memory with regard to the effect of labeling. Predictions regarding the hypothesis of the labeling effect were not preregistered. Experiment 1b was a replication with just one minor modification in the instruction and was not preregistered. We maintained the same preregistered analysis plan for both experiments.

Method

Participants

Fifty-seven students of the University of Zurich participated in this experiment. Only participants with German (or Swiss-German) mother tongue, aged between 18 and 35 years, and reporting normal color vision or corrected-to-normal visual acuity could take part in the experiment. Participants signed an informed consent prior to the study and were debriefed at the end. The experimental protocol was in accordance with the guidelines of the Institutional Review Board, and it did not require special approval.

The first 30 participants took part in Experiment 1a (M = 27.73, SD = 3.74, 23 women) and the next 27 participants were assigned to Experiment 1b (M = 23.19, SD = 3.56, 16 women). Six participants were excluded from Experiment 1a as they failed to follow the labeling instructions,[1] resulting in a final data set of 24 participants. Three participants were excluded from Experiment 1b,[2] resulting in a total of 24 participants. As detailed in our preregistration, we aimed to collect data of at least 30 participants in Experiment 1a, and we were going to adjust the sample size based on the evidence obtained for or against our hypotheses. The final sample size in these experiments was sufficient to provide substantial evidence to answer our research questions, hence we stopped data-collection as reported in the preregistration.

Materials

All experiments were programmed in MATLAB (2010b for Experiment 1; 2016b for Experiments 2 and 3) using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997). Nameable clip-art pictures served as stimuli objects, which were taken from Sutterer and Awh (2016). The objects were colored in one of 360 colors that varied along a continuous color wheel (Zhang & Luck, 2008), defined in the CIELAB color space with L = 70, a = 20, b = 38, and a radius of 60. The colored objects were presented against a gray background (RGB 128 128 128). Participants saw each object once. The color-object combinations (hereafter referred as bindings) were randomly selected for every participant.

Procedure

Visual Working Memory Phase. Each visual working memory trial started with a 1000 ms fixation cross in white (RGB 255 255 255) in the center of the screen. Thereafter, a sequence of three objects was presented. Each object remained onscreen for 250 ms, followed by a 1,000-ms blank interobject interval, providing time for labeling (see Figure 1A). To investigate how labeling influences visual working memory and visual long-term memory we introduced three labeling conditions during the study phase: (a) label the color (e.g., “red”), (b) label the object (e.g., “heart”), or (c) suppression (e.g., “bababa”). These labeling instructions appeared at the beginning of each trial to remind participants of the current condition. Participants were asked to self-initiate each trial by pressing the space bar. They were further instructed to wear a headset and their verbal responses were recorded for offline check. The labeling conditions were completed in short blocks of 8 trials, and blocks of different conditions alternated (e.g., suppression-color-object-suppression-color-object). The order of the conditions was counterbalanced across participants. Each condition contained three practice trials and 32 experimental trials. The practice trials were completed right before the first block of this condition. Overall, there were 105 objects per condition (including practice trials), and 315 objects in total.

In the visual working memory test, all three objects were tested in random order (see Figure 1B). The memory test phase was initiated by the presentation of a dark-gray wheel (RGB 96 96 96) around the tested object, which was presented in light gray (RGB 160 160 160). Once participants started moving the mouse along the gray wheel, the color of the probe changed. Participants were asked to adjust the color of the probe to the one they remembered for this object. Once participants right-clicked on the mouse, their color selection was registered, and the next object was presented.

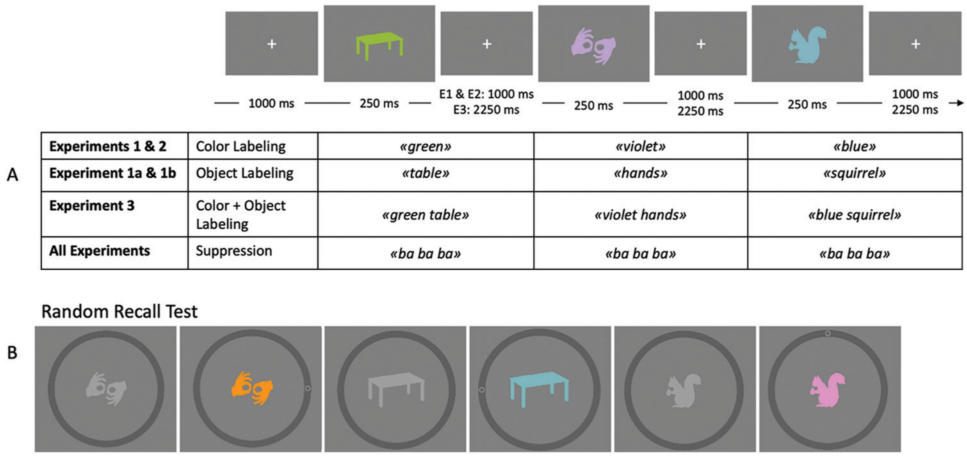
Visual Long-Term Memory Phase. At the end of the visual working memory phase, participants were instructed to leave the experimental room and take a short break for about 5 minutes. During the break, they were offered some sweets (e.g., chocolate). After the break, participants underwent the visual long-term memory test phase. This test phase matched the procedure of the visual working memory test. In Experiment 1a, participants were not aware of the visual long-term memory test, and hence the delayed test came as a surprise. In contrast, participants in Experiment 1b were informed prior to the start of the experiment that they would

[1] Four of these participants did not follow the instruction to switch between labeling conditions on several occasions and remained labeling the wrong condition for the entire block (e.g., they continued labeling the color instead of the object), and two did not label at all.

[2] One participant verbalized only on some trials, one participant confused the labeling conditions, and one participant labeled the fixation cross instead of the objects.

Figure 1

Illustration of the Flow of Events in the Trials of All Experiments Reported Here



Note. Panel A exemplifies the flow of one trial with examples of the actual objects used for all experiments. Below each object, the applied labeling conditions are illustrated. Panel B shows the random recall test procedure of this visual working memory trial. Participants first saw a probe in gray. Once participants moved the mouse along the wheel the object's color changed. For visual long-term memory, all objects were tested in the same manner. See the online article for the color version of this figure.

have to recall all of the presented objects at a second stage of the experiment, and they were encouraged to try to retain the objects for a longer duration in memory. In both experiments, participants were tested for all the objects from the visual working memory phase, excluding the practice trials. In total, 288 objects were tested in the visual long-term memory phase, 96 from each labeling condition.

Data Analysis

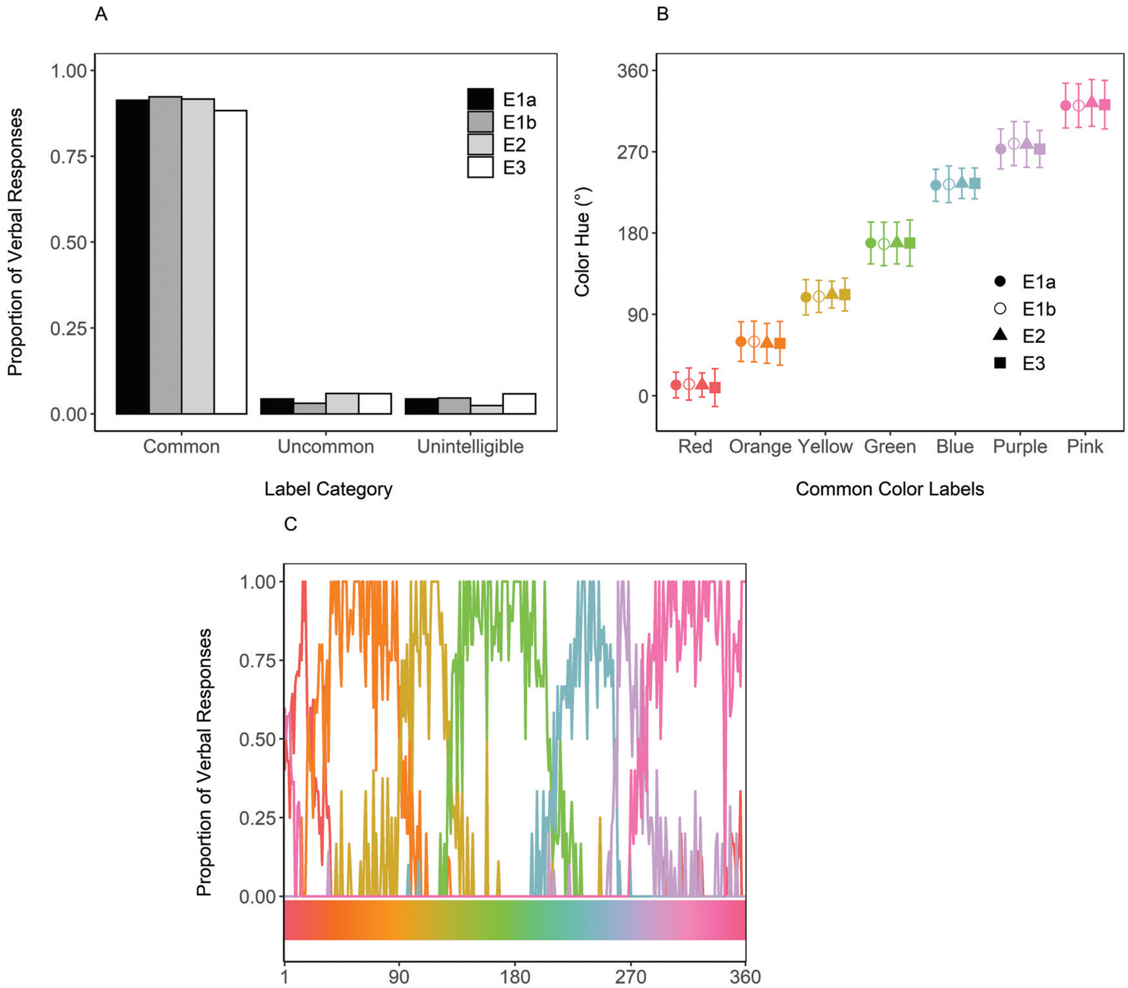
Verbal Labeling Output. We recorded the verbal responses during the study phase. Color labeling responses were coded to assess the variety of labels applied to the colors, and to estimate the color range to which these labels referred to in each experiment. This information was then used to inform our mixture modeling about participants' color categories in each experiment, following the procedure used by Souza and Skóra (2017).

Participants used a total of 20 different color labels in Experiments 1a and 1b, 50 in Experiment 2, and 76 in Experiment 3. Similarly to Souza and Skóra (2017), the majority of the color labels belonged to a set of basic color categories (e.g., red, orange, yellow, green, blue, purple, and pink) across all our reported experiments. Figure 2A shows the proportion of verbal responses that fell within these seven color categories (hereafter referred here as common category), as opposed to the usage of more uncommon labels (e.g., turquoise, yellow-green, dark orange, blueish), or unintelligible responses. This figure shows that although various labels were used overall, these uncommon responses were of very low frequency. Figure 2C presents the proportion of times that one of the seven basic color labels was used (across all participants) to refer to the 360 colors in the color wheel. This led to seven bell-shaped distributions across the continuous color space. The bell-shape of these distributions resembles a normal distribution, and hence we fitted a normal distribution for circular space (e.g., a von Mises distribution) to these data. The von Mises distribution is described by the mean and the standard deviation. These parameters can be taken to define the center of the color category and the variance around it. Figure 2B shows the center of each color category (dot) and the standard deviation of the color categories as estimated by the von Mises fitted to the verbal responses in each experiment.

Recall. Recall was assessed by calculating the deviation between the given response and the true color value of the studied object in degrees, ranging from +180 to −180 °. The absolute value of the deviation can be taken as a model-free index of performance, which we will refer here to as recall error. Our first set of analyses focused on differences between labeling conditions with regards to recall error in the visual working memory and visual long-term memory tests. We conducted Bayesian Inference statistics because this approach is known to have several statistical advantages over frequentist statistics that rely on p-value significance testing. For example, p values have the tendency to overstate evidence in favor of the alternative hypothesis (Wetzels et al., 2011). In contrast to p values, Bayesian inference quantifies the evidence for one hypothesis over the other. One commonly employed measure is the Bayes Factor (BF). The BF is the strength of evidence for one hypothesis (e.g., the alternative) over another hypothesis (e.g., the Null), given the observed data. The advantage of a Bayesian approach is that one can gauge evidence for the alternative and for the null hypothesis. A BF_{10} (e.g., the likelihood of the alternative hypothesis, H1, over the null hypothesis, H0) above 1 yields evidence in support of H1, whereas a BF_{10} below 1 provides evidence in support of H0. BFs should be interpreted as a continuous index of the strength of evidence in the data in support of one model over the other and provides the factor by which the ratio of our prior beliefs should be updated in light of the data. For example, a $BF_{10} = 10$ indicates that the alternative hypothesis is 10 times more likely than the null hypothesis, given the data. Usually, BFs > 3 are considered as providing substantial evidence for one hypothesis over the other, whereas a BF ≥ 10 is

Figure 2

Analysis of the Color Labels Used by the Participants Across All Experiments



Note. Panel A shows the proportion of color labels grouped by the common, uncommon, and unintelligible label categories for Experiments 1–3. Panel B shows the average color for which a given label was assigned and the standard deviation of colors to which the label was applied. These parameters were estimated by a von Mises fitted to the distribution of color label responses over the color space in all experiments. Panel C shows the proportion of times one of the seven common color labels was used to refer to a given color on the wheel (as shown in the *x* axis) in Experiment 1a. A proportion of 1 indicates that the *x* color on the wheel was labeled with the same label by all participants. The lower the proportion, the less often participants used that label to refer to the given color. Each color term is represented by the line with its prototypical color. E = Experiment. See the online article for the color version of this figure.

usually considered as strong evidence. We computed the BFs as stated in Rouder et al. (2012) using the default settings of the BayesFactor package (Morey & Rouder, 2015) implemented in R (R Core Team, 2014).

Experiment E1a and E1b were within-subject designs with 2 (memory test: visual working memory, visual long-term memory) × 3 (labeling condition: color, object, suppression) factors. These two factors were set as fixed predictors in the Bayesian ANOVA and the subject factor was treated as random effect. To compute a BF, the believed probabilities of the parameter distributions, also known as a priori beliefs or priors, need to be set judiciously and computationally convenient (Rouder et al., 2012). The Bayes Factor package provides three default priors that are within a reasonable range. Here, the BFs were computed with the most conservative default prior of $\sqrt{2}/2$. The chosen prior reflects our beliefs about the likelihood of an effect in our experiment. Rouder et al. (2017) showed that the prior specification matters, but it does not greatly change the evidence within a reasonable range of prior specifications, such as the range between .2 and 1 (which is within the range of our prior specification). The higher the BF, the less influential the prior is.

In the preregistrations we stated that we aimed to report BFs ≥ 10 for or against the alternative hypothesis for the main effects and the interactions of interest in the model, which is usually considered as strong evidence.

Categorical-Continuous Mixture Modeling. We modeled the responses in our task using the Bayesian hierarchical categorical-continuous mixture model of Hardman et al. (2017). The model assumes that responses are either informed by memory (P^M) or reflect guessing ($1 - P^M$). Responses informed by memory could reflect continuous (P^O) or categorical ($1 - P^O$) information about the visual stimulus. Continuous information allows for a fine-grained response that varies linearly with the studied feature. The continuous response can be more or less fine-grained—which reflects the continuous imprecision (σ^O) of the memory representation. In contrast, categorical responses cluster around some canonical values (the category mean) along the feature space. The model further assumes two sources of guessing: guessing could either be categorical, when participants randomly guess prototypical colors, captured by the parameter P^{AG}, or continuous, when guesses are uniformly distributed along the feature space ($1 - P^{AG}$). In this mixture model, every category has a mean and standard deviation, which can be estimated freely by the model if no prior knowledge about the participants' categories is given. In the following experiments, we fixed the category means using the information extracted from the labeling responses (see Figure 2B), similarly to the approach used by Souza and Skóra (2017).[3] Further parameters of the model are the category imprecision (how precise is the categorical response) and the categorical selectivity which estimates how selectively colors are assigned to a category.

For all analysis reported in this article, we fitted the between-item model of the CatContModel package (Hardman, 2016) implemented in R. The between-item model variant assumes that both categorical and continuous information relative to a stimulus can be held in memory at the same time. At the point of response selection, however, the response is based on either the categorical or the continuous information, but not both. This model variant has previously been reported to have better model fit (Hardman et al., 2017; Souza & Skóra, 2017) than the alternative variant assuming that responses reflect a combination of both continuous and categorical information. Hierarchical models view the parameters of individual participants in a given condition as samples from a population-level distribution. The parameter values and distributional probabilities were determined through Markov chain Monte Carlo (MCMC) sampling techniques.

For each experiment, we fitted a model that allowed the three main parameters P^M, P^O, and σ^O in the model to vary across experimental conditions. Then, we assessed the posterior estimates of the parameters of the model with regards to the effects of our manipulations. Our main interest was to assess how labeling changed the probability of responses informed by categorical as opposed to continuous information, and the continuous imprecision of the memory representation across both the visual working memory test and the visual long-term memory test. To assess the reliance on continuous information, P^M needed to be multiplied by P^O. To assess reliance on categorical information the equation is as follows: $P^M \times (1 - P^O)$. The continuous imprecision parameter (σ^O) was used as outputted by the model.[4]

Results

Recall Error

In the preregistration we mentioned to check the residuals of recall error for the assumption of a normal distribution by looking at the QQ plot of the residuals. To check the homogeneity of variance distribution for the recall error analysis, we calculated the variance of the mean recall error for every participant in every condition. The difference in variance in groups was below 4, which is the threshold for assumption violation (Tabachnick & Fidell, 2013). This information can be found on the OSF (https://osf.io/rkqth/).

The recall error as a function of labeling condition for the two memory tests is presented in Figure 3. Recall error was smaller in the visual working memory test than in the visual long-term memory test, reflecting better performance in the former. Compared with suppression, visual working memory performance improved when participants labeled the colors but decreased when participants labeled the objects. Labeling had no discernable effect on visual long-term memory performance.

In line with our preregistered analysis, we conducted a Bayesian ANOVA on the data of Experiments 1a and 1b. Table 2 presents the BFs of all tested models against the Null. The model with the highest BF against the Null is the best model. Our preregistered analysis was mainly concerned with the evidence for an interaction between labeling and memory test. The best model of the data in Experiments 1a and 1b included the effects of labeling condition, memory test, and their interaction. To assess the evidence for the inclusion of the interaction in the best model, we computed the ratio of the best model against the model with only the two main effects. As shown in Table 2, there was overwhelming evidence for the inclusion of the interaction between labeling condition and memory test in the best model of both experiments, indicating that labeling impacted visual working memory and visual long-term memory differently.

As a follow-up analysis on the interaction,[5] we assessed the impact of color and object labeling for the visual working memory and visual long-term memory test separately. We computed Bayesian *t* tests to compare both labeling conditions to the suppression condition. For visual working memory, the difference between color labeling and suppression yielded a $BF_{10} = 36.04$ in Experiment 1a and a $BF_{10} = 1.12 \times 10^3$ in Experiment 1b (both combined, $BF_{10} = 1.03 \times 10^5$), indicating strong support for a color labeling benefit. In contrast, the difference between object labeling and suppression yielded overwhelming evidence for an object labeling cost in both experiments (Exp. 1a: $BF_{10} = 3.54 \times 10^3$; Exp. 1b: $BF_{10}= 1.31 \times 10^6$; both combined, $BF_{10}= 2.92 \times 10^{10}$). For visual long-term memory, there was ambiguous to substantial evidence for the absence of a color labeling benefit—Exp. 1a: $BF_{10} = .91$ ($BF_{01} = 1.10$); Exp. 1b: $BF_{10} = .22$ ($BF_{01} = 4.64$); both combined, $BF_{10} = .36$. Likewise, there was ambiguous to substantial evidence against an object labeling cost in visual

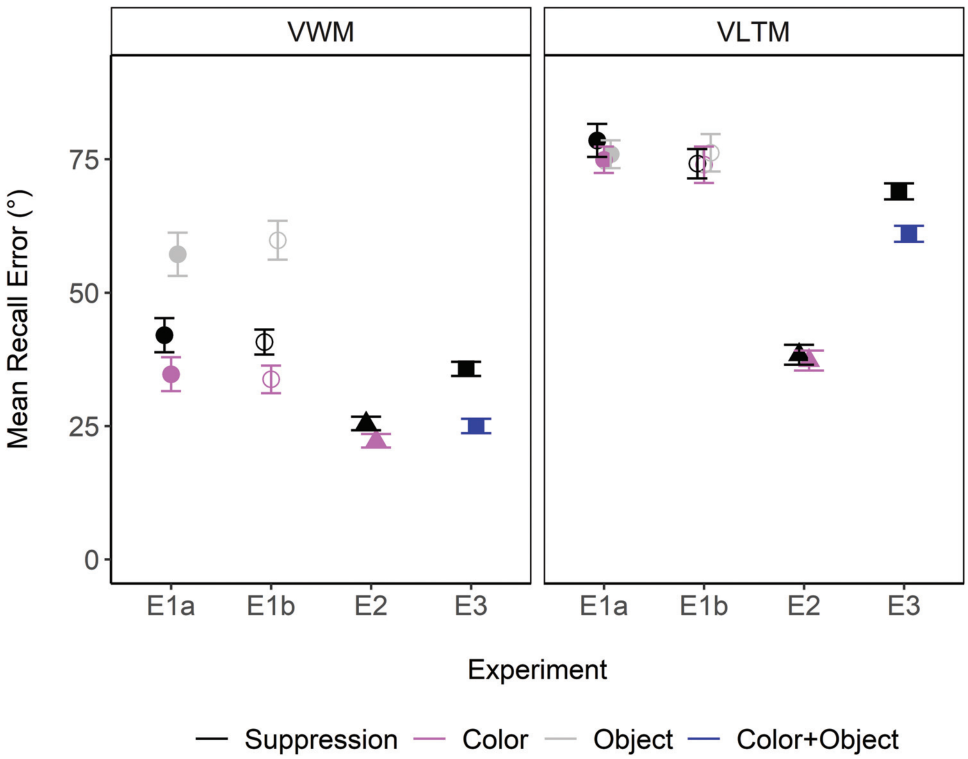
[3] We also fitted the model allowing free estimates of the color categories across all experiments, which can be found on the OSF (https://osf.io/rkqth/), and the results of these model were fairly in line with ones reported here (but see Experiment 3).

[4] In the preregistration, we mentioned that we would transform these values into the commonly used capacity K (Cowan, 2001), which requires multiplying the parameters by memory set-size. This produces, however, a very different scale range for visual working memory (zero to three items) and visual long-term memory (zero to hundreds of items). We decided therefore to keep parameters in the scale from zero to one for both memory systems. This decision is inconsequential for the assessment of the presence of effects.

[5] This set of analyses was not preregistered.

Figure 3

Mean Recall Error in Degrees Across All Experiments for the Visual Working Memory and Visual Long-Term Memory Tests



Note. The mean error in visual working memory for Experiment 2 averaged across the three repetitions of the same object. Error bars represent the 95% within-subjects confidence interval. VWM = visual working memory; VLTM = visual long-term memory; E = Experiment. See the online article for the color version of this figure.

long-term memory—Exp. 1a: $BF_{10} = .43$ ($BF_{01} = 2.30$); Exp. 1b: $BF_{10} = .33$ ($BF_{01} = 3.04$); both combined, $BF_{10} = .16$.

In sum, these results indicate that the color labeling benefit and object labeling cost found in visual working memory were no longer credible when memory was tested over a delay. Overt verbal labeling clearly affects visual working memory but seems to neither benefit nor harm visual long-term memory—with the latter being more evident in Experiment 1b, in which participants were aware of the upcoming visual long-term memory test.

Categorical-Continuous Mixture Modeling

To investigate how labeling affected the storage of categorical and continuous information in visual memory, we submitted our data to mixture modeling. We modeled the data of all participants and conditions simultaneously. We allowed the three main parameters in the model (namely P^M, P^O, and σ^O) to be affected by the two within-subjects predictors of labeling condition (suppression, color labeling, or object labeling) and memory test (visual working memory vs. visual long-term memory). Each model was restrained to a maximum of seven color categories, with their means taken from the verbal outputs (as shown in Figure 2B). For every model, we ran 10,000 iterations of which the first 1,000 were regarded as burn-in, leaving a total of 9,000 post burn-in iterations for analysis. The Appendix shows that the posterior estimates of all models across all experiments reproduced the actual data.

An aim of this study was to analyze how labeling would change categorical and or continuous information in memory. For this, we then calculated the amount of categorical and continuous information held in memory (categorical = $P^M \times (1 - P^O)$; continuous = $P^M \times P^O$). Figure 4 presents the mean group-level parameters (dots) and the 95% highest density interval (HDI; error-bars), obtained from the models in Experiments 1a and 1b. These values are also summarized in Table 3. These posteriors should be interpreted as follows: The mean represents the highest point of the posterior distribution and the HDI represents the range of values covering 95% of the posterior distribution. Hence, the HDI indicates the likely values of the parameter given the data. To estimate an effect for or against a verbal labeling benefit, one needs to compare the posteriors of, for example, the labeling condition against the posterior of the suppression condition. If the HDIs of these conditions do not overlap, it gives evidence for a labeling effect as performance between these two conditions substantially differs.

For visual working memory in Experiment 1a and 1b, there was a tendency of color labeling to increase total memory (P^M) in contrast to suppression (Figures 4A and 4E), but this effect was not fully credible as the HDIs of these conditions overlap. Color

Table 2

Relative Likelihood of Models With Different Fixed Effects Over the Null Model (BF_{10}) and Relative Likelihood of the Best Model (e.g., the One With Higher Likelihood Over the Null) Over the Alternative Model Specified in Each Row (BF_{Best}/BF_{Mrow})

Experiment	Model No.	Included Fixed Effects: Labeling condition	Memory test	Labeling × Test	BF_{10}	BF_{Best}/BF_{Mrow}
1a	**1**	✓	✓	✓	$\mathbf{1.17 \times 10^{47}}$	**1**
	2	✓	✓	—	2.68×10^{39}	4.38×10^{7}
	3	✓	—	—	3.06	3.83×10^{46}
	4	—	✓	—	6.94×10^{33}	1.69×10^{13}
1b	**1**	✓	✓	✓	$\mathbf{3.38 \times 10^{46}}$	**1**
	2	✓	✓	—	2.73×10^{37}	1.24×10^{9}
	3	✓	—	—	37.97	8.91×10^{44}
	4	—	✓	—	6.32×10^{28}	5.35×10^{17}
2	1	✓	✓	✓	3.36×10^{26}	3.05
	2	✓	✓	**—**	$\mathbf{1.02 \times 10^{27}}$	**1**
	3	✓	—	—	0.41	2.49×10^{27}
	4	—	✓	—	7.92×10^{26}	1.29
3	1	✓	✓	✓	2.76×10^{86}	1.78
	2	✓	✓	**—**	$\mathbf{4.91 \times 10^{86}}$	**1**
	3	✓	—	—	45.52	1.08×10^{85}
	4	—	✓	—	4.89×10^{69}	1.00×10^{17}

Note. ✓ = effect included in the model. The model with the highest BF against the Null (best model) is printed in bold.

labeling had no credible effect on the probability of retrieving categorical information (see Figures 4B and 4F). There was a tendency for labeling to increase continuous memory (Figures 4C and 4G) and reduce memory imprecision (Figures 4D and 4H) in comparison to suppression, but this effect was again not credible. In contrast, object labeling led to a credible reduction of total memory and on the probability of retrieving categorical information compared with suppression. Object labeling also had credible costs for continuous memory: In Experiment 1a, this was revealed by a reduction in continuous precision (Figure 4D), whereas in Experiment 1b this translated into a lower probability of retrieving continuous representations (Figure 4G).

For visual long-term memory, the HDIs of all labeling conditions overlap across all three parameters in both experiments, showing no credible effects of labeling.

Discussion

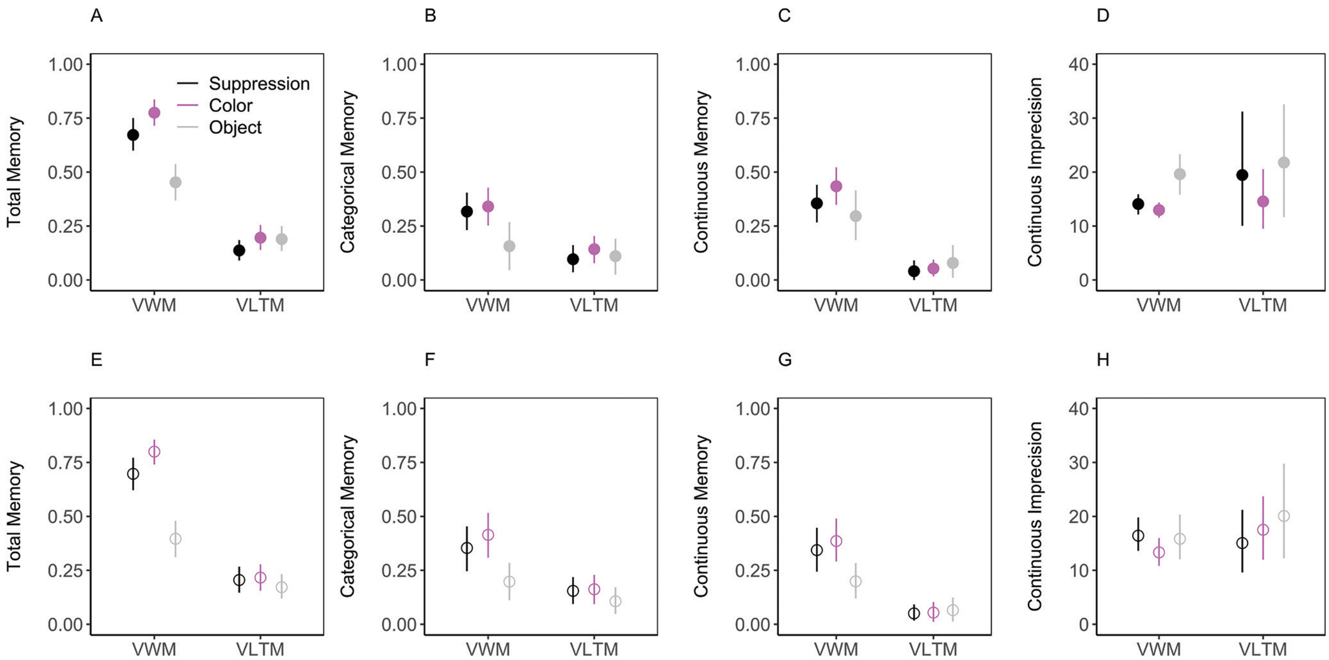
In both experiments, labeling the color of the colored objects was helpful for the retention of this feature in visual working memory compared with a condition in which labeling was inhibited with articulatory suppression—as revealed by the recall error measure. With regard to mixture modeling, color labeling tended to increase the accessibility of representations overall and tended to improve memory precision, but in this series of experiments these effects were not credible. These results are in line with the ones of Souza and Skóra (2017) in which color labeling was found to aid the maintenance of color representations in visual working memory, extending it to a paradigm in which participants maintained color-object bindings. Furthermore, Experiment 1 showed that labeling another feature of the object (its shape) was detrimental to the retention of color information in visual working memory, reducing both categorical and continuous information. This happened although object-labeling provided a unique cue to the studied object (given that each object was only presented once). These findings rule out several hypotheses of the labeling effect for visual working memory (see Table 1), namely all hypotheses but Hypotheses 4 and 5: labeling the colors seems to activate categorical representations that boost memory for color, whereas labeling other features directs attention away from this feature yielding a cost. Altogether our findings indicate that labeling is only beneficial for visual working memory if it provides categorical information about the relevant feature of the object.

Critical to our main research question, the delayed test showed that the visual working memory effects of labeling were short-lived. In line with the results of Kelly and Heit (2017), labeling did not affect visual long-term memory irrespective of whether participants were aware (Experiment 1b) or not (Experiment 1a) of the upcoming visual long-term memory test. This suggests that labeling impacts visual representations differently over the short and long term.

There is one caveat, though. Overall recall error in the visual long-term memory test was around 75 degrees. Given that chance performance in this task is associated with a recall error close to 90 degrees, the lack of a labeling effect might be related to poor visual long-term memory learning overall. Simple knowledge about the upcoming visual long-term memory test was not sufficient to yield better performance in this task, given that visual long-term memory performance was similar across the experimental versions in which the delayed test was a surprise (Experiment 1a) or was announced at the beginning of the study (Experiment 1b). It is possible that labeling does foster learning in visual long-term memory, but the number of objects learned (315 in total) and the slim opportunities to commit this information to memory (single study opportunity) precluded us from observing this beneficial effect. The goal of Experiment 2 was to address this possibility.

Figure 4

Parameter Estimates of Categorical-Continuous Modeling for Experiments 1a (A–D) and Experiment 1b (E–H)



Note. Dots depict the mean of the posterior distributions and the error bars depict the 95% highest density interval (HDI). The two left columns show estimates of probability of retrieving total information, which is defined by the probability to retrieve categorical information (second column from left), and continuous representations (third from left), and the right columns shows estimates of continuous memory imprecision. VWM = visual working memory; VLTM = visual long-term memory. See the online article for the color version of this figure.

Experiment 2

In Experiment 1 participants had a single opportunity to study a color-object binding, and they were only tested on this binding once in visual working memory and once in visual long-term memory. The objects were studied in conditions that differed in the opportunity to label certain binding features: Participants labeled the color, the object, or they repeated "bababa" aloud to prevent labeling. Color labeling was beneficial and object labeling was detrimental for the storage of the color-object bindings in visual working memory compared with suppression. Notwithstanding, all conditions yielded the same level of visual long-term memory performance.

These results point to a dissociation between learning over the short and long term. Conditions that fostered and hampered visual working memory had no impact on the retention of representations in visual long-term memory. This is not in line with studies suggesting a link between memory over the short and long term (Biderman et al., 2019; Brady et al., 2013; Oberauer et al., 2017). Following up on this issue, Experiment 2 addressed the possibility that labeling did not affect visual long-term memory due to the limited opportunities to learn the color-object associations. Previous studies have shown that long-term learning is fostered by repeated testing of memory compared with restudying (Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger & Pyc, 2012). Recently, this *testing effect* was also found to occur for visual long-term memory (Sutterer & Awh, 2016). Sutterer and Awh (2016) presented participants with colored objects for study for a total of 400 images. For half of the images, participants restudied the color a second time; whereas for the other half, they practiced recalling their color. In a final test, participants reproduced the colors of all objects (visual long-term memory test). Visual long-term memory performance for the tested objects was higher than for the restudied objects. Along with the testing effect, it has been shown that repeated presentation of information also increases visual working memory performance (Couture & Tremblay, 2006). This repetition effect, also known as the *Hebb effect*, consists of the observation of better recall for memory lists as a direct function of the number of times the list was repeated during the course of the experiment. Recently, Miner et al. (2020) showed that visual long-term memory for colored objects can be improved by item repetitions, reaching visual working memory levels if the objects are restudied eight times.

The aim of the present experiment was, first, to leverage the repetition and testing effects to increase visual long-term memory performance in the delayed test at the end of the study. Our second aim was to assess whether color labeling could foster long-term learning as reflected in the rate of learning over repetitions (e.g., during the visual working memory phase). To test for this, the color-object associations were repeated three times in a row (e.g., over three successive visual working memory trials). Our two main questions were whether the visual working memory improvement over repetitions (e.g., the learning rate) would be different across the labeling conditions and whether this would translate into different performance levels in the delayed recall test in the final visual long-term memory phase.

Table 3
Posterior Means and Highest Density Intervals (HDI) of the Mixture Model Parameters in All Experiments

Exp. + Condition + Repetition (R)	Visual working memory						Visual long-term memory					
	Categorical		Continuous		Continuous imprecision		Categorical		Continuous		Continuous imprecision	
	M	95 % HDI	*M*	95 % HDI	*M*	95 % HDI	*M*	95 % HDI	*M*	95 % HDI	*M*	95 % HDI
E1a Suppression	0.32	[0.23, 0.40]	0.36	[0.27, 0.44]	14.08	[12.16, 15.88]	0.10	[0.04, 0.16]	0.04	[1.53×10^{-6}, 0.09]	19.46	[10.04, 31.20]
E1a Color	0.34	[0.25, 0.43]	0.43	[0.35, 0.52]	12.96	[11.55, 14.37]	0.14	[0.08, 0.20]	0.05	[0.02, 0.09]	14.55	[9.51, 20.55]
E1a Object	0.16	[0.05, 0.27]	0.30	[0.19, 0.42]	19.63	[15.84, 23.32]	0.11	[0.03, 0.19]	0.08	[0.01, 0.16]	21.76	[11.66, 32.57]
E1b Suppression	0.35	[0.25, 0.45]	0.34	[0.24, 0.45]	16.42	[13.62, 19.79]	0.15	[0.09, 0.22]	0.05	[0.02, 0.09]	15.05	[9.60, 21.21]
E1b Color	0.41	[0.31, 0.52]	0.39	[0.29, 0.49]	13.32	[10.83, 15.97]	0.16	[0.09, 0.23]	0.05	[0.01, 0.10]	17.52	[11.98, 23.72]
E1b Object	0.20	[0.11, 0.28]	0.20	[0.12, 0.28]	15.84	[12.05, 20.33]	0.11	[0.05, 0.17]	0.07	[0.01, 0.12]	20.07	[12.23, 29.78]
E2 Suppression R1	0.18	[0.12, 0.24]	0.54	[0.46, 0.61]	17.19	[16.01, 18.50]						
E2 Color R1	0.28	[0.21, 0.35]	0.53	[0.45, 0.60]	13.04	[11.98, 14.03]						
E2 Suppression R2	0.29	[0.21, 0.37]	0.65	[0.56, 0.72]	13.09	[12.13, 14.21]						
E2 Color R2	0.29	[0.21, 0.35]	0.68	[0.59, 0.74]	11.66	[10.80, 12.56]						
E2 Suppression R3	0.22	[0.16, 0.29]	0.74	[0.67, 0.80]	12.62	[11.78, 13.49]						
E2 Color R3	0.27	[0.20, 0.34]	0.40	[0.63, 0.77]	11.31	[10.46, 12.13]						
E2 Suppression	0.28	[0.23, 0.34]	0.60	[0.54, 0.65]	13.14	[12.43, 13.83]	0.28	[0.20, 0.36]	0.43	[0.36, 0.51]	15.10	[13.78, 16.54]
E2 Color	0.29	[0.24, 0.35]	0.62	[0.57, 0.68]	11.57	[10.92, 12.21]	0.38	[0.30, 0.46]	0.35	[0.28, 0.43]	13.40	[11.82, 14.84]
E3 Suppression	0.41	[0.36, 0.46]	0.37	[0.31, 0.41]	13.82	[12.81, 14.72]	0.21	[0.17, 0.26]	0.06	[0.03, 0.09]	15.22	[11.63, 19.03]
E3 Color + Object	0.45	[0.40, 0.51]	0.47	[0.41, 0.52]	13.84	[12.99, 14.68]	0.32	[0.26, 0.37]	0.07	[0.05, 0.10]	12.40	[9.86, 15.24]

Note. ✓ = effect included in the model; E = Experiment. The model with the highest BF against the Null (best model) is printed in bold.

We predicted that visual working memory performance would increase across repetitions along with the creation of stronger visual long-term memory traces. Regarding the effects of labeling, we hoped to distinguish between two possible scenarios. One possibility is that labeling only helps over the short term as suggested in Experiment 1. If this is the case, we should observe a labeling benefit in visual working memory, but labeling should not (a) alter the rate of visual long-term memory learning over the repetitions and (b) it should not yield better recall in the delayed test. Another possibility is that with more opportunities to learn the color-object bindings, labeling would be beneficial both over the short and long term (e.g., with more learning over repetitions and better delayed recall). This would indicate that the long-term beneficial effect of labeling may be too weak to be observed in single-trial learning but does accumulate over repetitions. These hypotheses, the experimental design, and the analysis plan for Experiment 2 were preregistered and can be found at https://osf.io/tker5/.

To foreshadow our results, the color labeling benefit was yet constrained to visual working memory. We only found a beneficial effect of labeling on the very first exposure to the color-object binding. Over the course of the repetitions, the color labeling advantage vanished within visual working memory, and it was absent in the final visual long-term memory test. Together with Experiment 1, these results point toward a dissociation on the impact of verbal labeling for memory over the short and long term.

Method

Participants

In total, 60 participants (*M* = 23.38, *SD* = 3.89, 42 women) of the University of Zurich took part in this experiment, 58 of these participants had not taken part in an experiment reported here. Participants fulfilled the same criteria and were exposed to the same protocol as in Experiment 1a and 1b. Note that we started the experiment with a sample of 30 participants; however, as we obtained ambiguous evidence for the interaction of labeling and memory system (visual working memory vs. visual long-term memory), we increased our sample size until the maximum preregistered sample size was reached.

Materials

The same materials as in Experiment 1 were used. In total, 102 objects were chosen randomly for every participant out of the set of 315 objects used in Experiment 1. The color of the objects was randomly assigned and sampled from the same color wheel as in Experiment 1.

Procedure

Visual Working Memory Phase. The visual working memory phase of Experiment 2 followed the same procedure as in Experiment 1b with the following changes. First, Experiment 2 included only two conditions: color labeling and suppression; the object labeling condition was removed. The reason for this was that we wanted to focus on conditions that could improve memory. Second, each visual working memory trial was presented three times in a row. More specifically, a trial consisted of the sequential presentation of three color-object bindings, and in Experiment 2 the exact same color-object bindings were repeated over three consecutive trials. We thereby lowered the number of objects participants had to learn in contrast to Experiment 1. Third, the order of presentation of the colored objects varied for every trial repetition to ensure that participants learned the color-object binding (e.g., pink-mug; blue-shoe; green-bucket) and not the order of the colors (pink-blue-green). After every trial, a test phase followed where memory for the colors of the three objects was tested in random order. To simplify, these three trial repetitions are hereafter referred to as one miniblock.

The experiment was divided into six blocks consisting of five miniblocks each (three with each labeling condition). The manipulation of color labeling and suppression occurred across blocks, which alternated throughout the experiment. Presentation order of the blocks was counterbalanced across participants. In total, participants completed 90 experimental trials consisting of three repetitions of 30 unique sets of three memory objects. Participants learned 90 objects, 45 objects in the color labeling condition and 45 objects in the suppression condition. To familiarize participants with the task, they performed two practice miniblocks (six trials) of each labeling condition before the exposition to the first experimental block with that condition. The practice blocks were excluded from further analysis. As in Experiment 1b, participants were informed prior to the start of the experiment that they should aim to retain the objects for a longer duration and that they would be asked to recall them again at a second stage in the experiment.

Visual Long-Term Memory Phase. After the end of the visual working memory phase, participants completed a multiplication verification task for about 2 min. In this task, simple multiplications (e.g., 3 × 8 = 25?) were presented on screen, and participants indicated whether the result was correct or not by pressing the right-arrow key or the left-arrow key, respectively. In total, 40 multiplications were verified. The reason for imposing this task was to eliminate the effect of recency of presentation of the last visual working memory trials. Next, participants were tested on the colors of the 90 objects learned in the visual working memory phase in random order. The test was as described for the visual working memory phase.

Results

Learning Effect on Recall Error

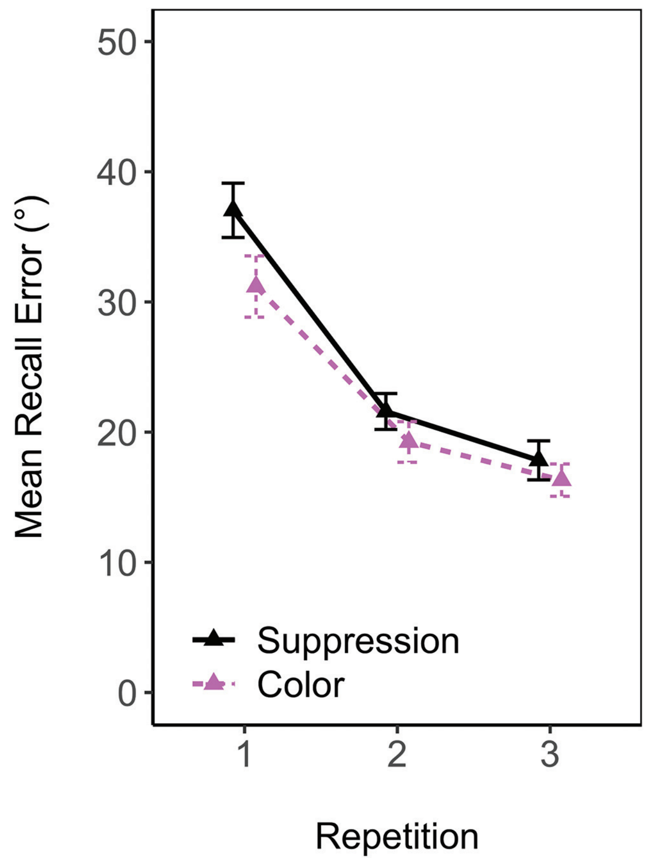
We first assessed the effect of labeling on learning over the three repetitions in the visual working memory task. Figure 5 shows the mean recall error across repetitions. A color labeling benefit is visible only in the very first exposure to the color-object binding.

Table 4 shows the analysis of the visual working memory test including the predictors of labeling condition (suppression vs. color labeling) and repetition (1 vs. 2 vs. 3). The best model of the data included all main effects and their interaction; however, there was ambiguous evidence for the inclusion of the interaction as a predictor in the best model even after collecting data of 60 participants.[6] We then followed up analyzing the effect of the interaction by conducting Bayesian ANOVAs between the labeling conditions for each repetition independently[7] (see Table 4). The comparison between the first and second repetition revealed ambiguous evidence against the inclusion of an interaction in contrast to the model with the two main effects. The best model for the comparison between the first and third repetition included the two main effects, but the exclusion of the interaction term was again ambiguous. The comparison of the second to the third repetition included both main effects and this model was substantially favored over the model including the interaction between the two predictors.

Finally, we conducted Bayesian *t*-tests contrasting the color labeling and suppression conditions for each repetition independently to estimate a potential color labeling benefit. For the very first presentation, there was strong evidence for a color labeling benefit (BF_{10} = 72.14). For the second and third presentations, however, there was no clear evidence for either the presence of absence of a color labeling effect (BF_{10} = .98/BF_{01} = 1.01; BF_{10} = .67/BF_{01} = 1.49).

Figure 5

Recall Error as a Function of Repetition and Labeling Condition in Experiment 2



Note. Error bars represent the 95% within-subjects confidence interval. See the online article for the color version of this figure.

Overall Recall Error

The recall error between labeling conditions and memory tests is presented in Figure 1. For this analysis, visual working memory performance reflects the average performance over the three repetitions. Performance was better in the visual working memory test than in the visual long-term memory test. Similar levels of performance were obtained for the color labeling and suppression conditions in both memory tests.

In the preregistration, we stated that we would analyze the data similarly to Experiment 1. The results of the Bayesian ANOVA are presented in Table 2. The best model included both main

[6] The interaction was similarly ambiguous for the sample size of 30 participants.

[7] This set of analysis was not preregistered because it was a follow-up on the interaction (https://osf.io/rkqth/).

Table 4

Relative Likelihood of Models With Different Fixed Effects Over the Null (BF_{10}) and Relative Likelihood of the Best Model (e.g., the One With Higher Likelihood Over the Null) Over the Alternative Model Specified in Each Row (BF_{Best}/BF_{Mrow}) for the Recall Error in the Visual Working Memory Phase of Experiment

		Included fixed effects				
Rep. (R)	Model No.	Labeling condition	Repetition	Labeling × Repetition	BF_{10}	BF_{Best}/BF_{Mrow}
All R	**1**	✓	✓	✓	$\mathbf{3.45 \times 10^{55}}$	**1**
	2	✓	✓	—	2.86×10^{55}	1.20
	3	✓	—	—	6.61	5.21×10^{54}
	4	—	✓	—	1.67×10^{52}	2.03×10^{3}
R: 1 vs. 2	1	✓	✓	✓	4.01×10^{28}	1.75
	2	✓	✓	—	$\mathbf{4.61 \times 10^{28}}$	**1**
	3	✓	—	—	7.37	6.25×10^{27}
	4	—	✓	—	9.20×10^{25}	501
R: 1 vs. 3	**1**	✓	✓	✓	$\mathbf{1.24 \times 10^{38}}$	**1**
	2	✓	✓	—	6.32×10^{37}	1.96
	3	✓	—	—	1.70	7.28×10^{37}
	4	—	✓	—	5.92×10^{35}	209
R: 2 vs. 3	1	✓	✓	✓	2.66×10^{4}	4.32
	2	✓	✓	—	$\mathbf{1.15 \times 10^{5}}$	**1**
	3	✓	—	—	5.14	1.24×10^{4}
	4	—	✓	—	1.30×10^{4}	8.82

Note. ✓ = effect included in the model. R = repetition.

effects. This model was preferred over the model including an interaction between labeling and memory system. Furthermore, comparison of the best model against the model with only the effect of memory revealed ambiguous evidence for the inclusion of labeling condition as a predictor. We then followed up analyzing the labeling effect by conducting Bayesian *t*-tests between color labeling and suppression for visual working memory and visual long-term memory test separately. There was a clear labeling benefit for visual working memory, $BF_{10} = 45.26$. In contrast, there was evidence for the absence of a labeling benefit in visual long-term memory, $BF_{10} = .19$ ($BF_{01} = 5.26$).

Learning Effect on Categorical-Continuous Mixture Model Parameters

In the first model, we assessed the impact of labeling and repeated presentation on visual working memory. The model included the factor labeling condition (suppression vs. color labeling) and repetition (1 vs. 2 vs. 3). We fitted the model with 10,0000 iterations from which we discarded 1,000 iterations as burn-in, resulting in 9,000 post burn-in iterations for analysis.[8]

This model's posterior means and HDIs for our condition of interest can be found in Figure 6 and their respective values in Table 3. The probability of retrieving categorical representations (Figure 6A) was not credibly different between labeling conditions in any of the three repetitions. Repetition increased categorical memory from presentation 1 to 2, but not further with the third repetition. The probability of retrieving continuous representations (Figure 6B) was generally not affected by labeling, but it increased monotonically with repetition. Lastly, Figure 6C clearly shows that labeling led to more precise continuous memory on the first repetition in contrast to suppression. This boost in continuous precision, however, was substantially reduced in the subsequent repetitions, and it was no longer fully credible. One should note that repetition (particularly from first to second presentation) credibly reduced memory imprecision.

Visual Working Memory Versus Visual Long-Term Memory through the Categorical-Continuous Mixture model Parameters

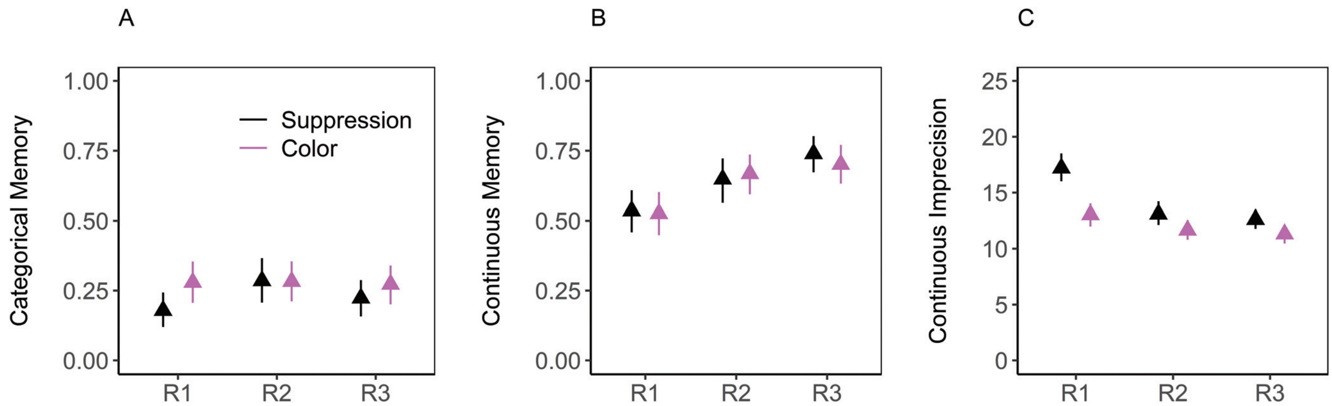
We then assessed the impact of labeling (color vs. suppression) and the two types of memory tests (visual working memory vs. visual long-term memory) on the parameters of the categorical and continuous memory mixture model. We again used the color category constraints of the verbal outputs and set the number of categories to seven. The model fit consisted of 10,0000 iterations from which we discarded 1,000 iterations as burn-in.

The posterior means and HDIs for our conditions of interest can be found in Figure 7, whereas the summaries of the estimates are presented in Table 3. Figure 7A shows that categorical memory did not differ between labeling conditions for visual working memory, but it was somewhat higher for color labeling compared with suppression in visual long-term memory, although this effect was not credible. Figure 7B shows that continuous memory was again not affected in visual working memory, but for visual long-term memory it was somewhat reduced (again not credibly) by color labeling. Last, continuous imprecision was credibly smaller for color labeling than suppression in visual working memory. There was a small tendency that this was also the case for visual long-term memory, but this was once again not credible.

[8] We also modeled the data without constraining the color categories. The results of this analyses can be found in the OSF (https://osf.io/rkqth/). In general, this analysis yielded a similar pattern to the one reported here (https://osf.io/rkqth/).

Figure 6

Mixture Model Parameters (Mean and 95% Highest Density Interval [HDI]) for the Visual Working Memory Data of Experiment 2



Note. R = repetition. Panel A shows probability of retrieving categorical representations, Panel B shows probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision. See the online article for the color version of this figure.

Discussion

Similar to Experiment 1, Experiment 2 showed a facilitative effect of verbal labeling that was restricted only to visual working memory despite our efforts to improve long-term learning. Again, the pattern of visual working memory benefits we observed was in line with the activation of categorical visual long-term memory hypothesis: Labeling boosted continuous memory, here reflected in a reduction of memory imprecision. This is consistent with prior findings in which either the number of continuous representations credibly increased or continuous memory imprecision was credibly reduced, but not both (Souza & Skóra, 2017).

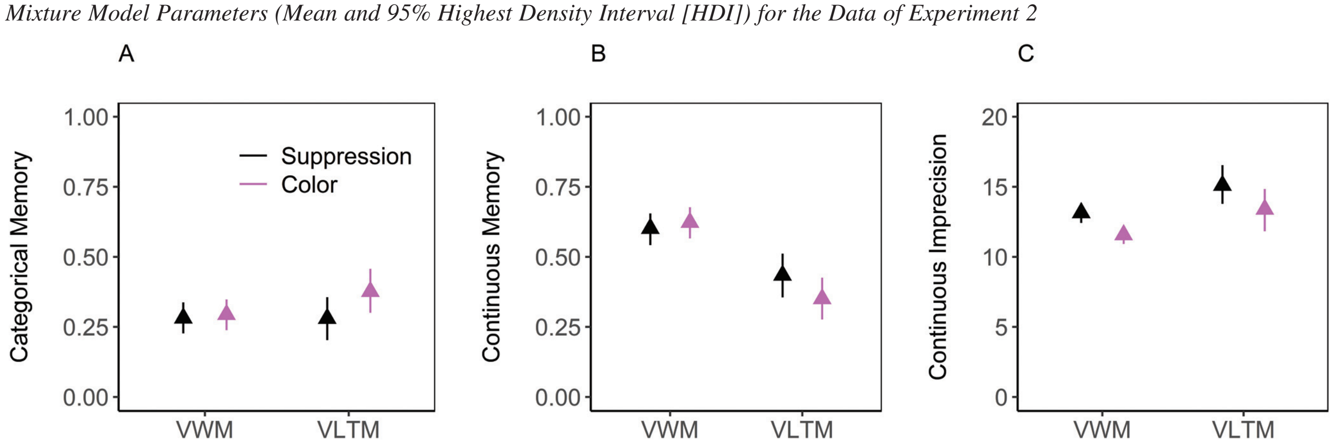
In Experiment 2, we repeated the presentation of the memoranda three times, and this improved performance overall over the short and long term (see also Miner et al., 2020). The repetition benefit was reflected in all parameters of the mixture model: The number of categorical and continuous representations stored increased, and memory imprecision decreased. Critically, however, labeling the colors did not facilitate learning: improvements over the repetitions were not influenced by labeling and neither was performance in the final delayed test. This addresses one concern raised in Experiment 1, namely, that the beneficial effect of labeling was not detected due to low long-term learning. So far, our results show that labeling the colors of visual objects boosts visual working memory, but not visual long-term memory. In our last experiment, we assessed whether this result generalizes to conditions in which both the color and the object features are labeled concurrently.

Experiment 3

The previous experiments implemented labeling conditions where either the color or the object was labeled, but not both simultaneously. In Experiment 1, labeling the color was beneficial, whereas labeling the object was detrimental to visual working memory. This raises the question whether labeling both features would yield any benefit at all. The main aim of Experiment 3 therefore was to assess whether labeling the association between

Figure 7

Mixture Model Parameters (Mean and 95% Highest Density Interval [HDI]) for the Data of Experiment 2



Note. Panel A shows probability of retrieving categorical representations, Panel B shows probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision. VWM = visual working memory; VLTM = visual long-term memory. See the online article for the color version of this figure.

the color and the object could be beneficial over the short and the long term. With regards to visual working memory, there are three different possible scenarios: (a) The beneficial effect of color labeling is also observed when, in addition to color, the object is labeled; (b) Because labeling the color is beneficial, but labeling the object is costly, these two effects cancel each other out and no effect is observed when both the color and the object are labeled; (c) The impairment of object labeling in visual working memory prevails when labeling both the object and color. We again tested whether the effects observed over the short term would be retained when memory is tested after a delay (visual long-term memory test). These hypotheses, the experimental design, and the analysis plan were preregistered and can be found at: https://osf.io/k3nsc/.

To foreshadow our results, labeling the color-object association was beneficial in visual working memory and, for the first time, we found evidence that this benefit remained in visual long-term memory. This indicates that labeling in visual working memory only translates into better visual long-term memory when the binding, in this case both the object and its color, are labeled concurrently.

Method

Participants

In total, 60 new participants (M = 24.47, SD = 4.30, 45 women) of the University of Zurich were tested under the same constraints as in Experiment 2. Data of two participants were excluded because they did not comply with the labeling instructions (one did not label at all, and one labeled only the colors on more than 70% of the occasions). We again note that, in line with our preregistration, we first tested 30 participants. As evidence for the effect of labeling across memory systems was in the ambiguous range, we doubled the sample size following our registered plan.

Materials

In total, 312 objects were presented to every participant. Colors were assigned randomly to each of the objects.

Procedure

Visual Working Memory Phase. The visual working memory phase of Experiment 3 followed the same procedure as in Experiment 1 with the following exceptions: First, Experiment 3 included two labeling conditions: color + object labeling versus suppression. In the color + object labeling condition, participants were instructed to overtly label the presented color and the object (e.g., "blue heart"), whereas in the suppression condition participants were instructed to articulate "bababa" aloud. Second, in this experiment every trial consisted of the sequential presentation of three objects, with each object being onscreen for 250 ms, followed by a 2,250-ms interstimulus blank interval. The interstimulus interval was increased to accommodate for the fact that labeling the binding takes longer than labeling only one single aspect of the stimulus. Accordingly, the same amount of time was provided for the suppression condition. The color + object labeling and suppression trials alternated every 10 trials throughout the experiment, and the order of labeling conditions was counterbalanced across participants. The experiment consisted of 104 trials, 52 for each labeling condition, of which the first two trials in each block were regarded as practice trials, resulting in 50 experimental trials in each condition. As in Experiment 1b, participants were informed that they needed to recall the objects at a later point in time and were asked to try to remember them for a longer period.

Visual Long-Term Memory Phase. After the visual working memory task participants took a short break, in which they left the experimental room and were offered some sweets (e.g., chocolate). Then, participants were tested again on the colors of the 300 objects (12 objects from the practice trials not included) learned in the visual working memory phase in random order.

Results

Recall Error

The mean recall error for each memory test and labeling condition are visualized in Figure 1. Visual inspection clearly shows that performance in visual working memory is better than for visual long-term memory, in line with all of the previous experiments. There is a benefit for labeling the color+object association in visual working memory compared with saying "bababa." For the first time in our series of experiments, there was a labeling benefit in visual long-term memory, as the recall error in the color + object labeling condition was smaller than in the suppression condition.

We preregistered to analyze the data in accordance with the previous experiments. The results of the Bayesian ANOVA are presented in Table 2. The best model of the data included the main effects of labeling and memory test. However, there was ambiguous evidence for excluding the interaction of labeling and memory test from the best model, even after we collected data of 60 participants. Bayesian t tests yielded evidence for a clear labeling effect within visual working memory, $BF_{10} = 2.30 \times 10^{13}$, and also within visual long-term memory, $BF_{10} = 1.98 \times 10^{7}$. Hence, the ambiguous interaction is not attributable to labeling not being beneficial over the long term, but it seems to relate to ambiguous evidence regarding whether this benefit is of the same size in visual working memory and visual long-term memory. Regardless of whether this benefit is of the same size or not, the critical point is that Experiment 3 showed, for the first time, evidence for a labeling benefit in episodic visual long-term memory. This suggests that a long-lasting labeling benefit is constrained to conditions in which bindings are labeled.

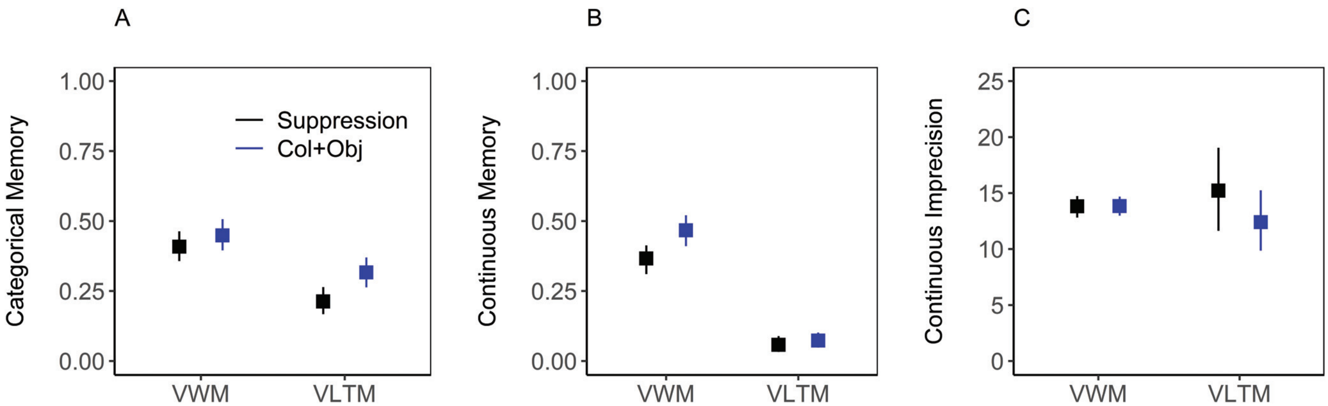
Categorical-Continuous Mixture Model

As in the previous two experiments, we assessed the impact of labeling and the two memory tests on categorical and continuous memory along with continuous imprecision. We again used the color category constraints of the verbal outputs and set the number of categories to seven. The model included the factor labeling condition (suppression vs. color+object labeling) and memory (visual working memory vs. visual long-term memory). The model fit consisted of 10,0000 iterations from which we discarded 1,000 iterations as burn-in.

The posterior means and HDIs for our conditions of interest can be found in Figure 8 and the summary of the estimates in Table 3. Figure 8A shows that categorical memory was somewhat higher in the color+object labeling condition in comparison to suppression in visual working memory, but this increase was not credible. The

Figure 8

Estimated Mixture model Parameters (Mean and 95% Highest Density Interval [HDI]) for the Data of Experiment 3



Note. Panel A shows the probability of retrieval of categorical representations, Panel B shows the probability of retrieval of continuous representations, and Panel C shows continuous memory imprecision. VWM = visual working memory; VLTM = visual long-term memory. See the online article for the color version of this figure.

same pattern is visible for visual long-term memory, but here the increase in categorical memory was credible. Continuous memory (Figure 8B), in contrast, was only credibly higher for the labeling than the suppression condition for visual working memory, but not for visual long-term memory. Last, continuous imprecision (Figure 8C) did not show a labeling effect, neither in visual working memory nor in visual long-term memory. For visual long-term memory, there is a small but not credible tendency of a decrease due to labeling.

We also fitted the model allowing free estimates of the color categories and the results were fairly in line with the ones of the model with the constrained color categories, except that, for continuous memory, the labeling benefit was smaller and not credible.

Discussion

Experiment 3 replicated the finding of a labeling benefit in visual working memory in contrast to a suppression condition. This time, labeling was not only helpful when the color itself was labeled, but rather when the color and object binding was labeled extending the scope of the labeling effect in visual working memory. This stands in contrast to the fact that when only the object was labeled in Experiment 1, it led to the forgetting of the color (as if it led to the filtering of this information). However, when the object was labeled alongside the color, it no longer competed with the relevant color information, and both features could be stored.

For the first time in our series of experiments, we could show that a labeling benefit in visual working memory was translated into better visual long-term memory. When analyzed with the mixture model, the data showed that the sources of the labeling benefit were different between visual working memory and visual long-term memory: Replicating our previous experiments and Souza and Skóra (2017), labeling improved storage of continuous representations in visual working memory. In contrast, for visual long-term memory, the beneficial effect of labeling was mainly due to categorical representations, with no credible changes to continuous memory.

General Discussion

Across three experiments, we found a labeling benefit in visual working memory when participants labeled the color of a colored object. In Experiment 1 and 2 we showed that color labeling benefited continuous color recall in visual working memory compared with a suppression condition. Additionally, Experiment 1 showed that labeling the object's identity yielded a cost to the retention of the object's color compared with suppression. This indicates that color information is lost when participants label another feature of the visual object (e.g., its shape). When both the color and object's identity were labeled concurrently though (Experiment 3), there was only a labeling benefit in visual working memory. These findings extend previous results by confirming that labeling affects the storage of visual information in visual working memory (Souza & Skóra, 2017; see also Forsberg et al., 2020; Souza et al., 2021). Our results show that labeling adds information to the visual features stored in visual working memory, and this can lead to augmented retention of the labeled feature, even if this may come at the expense of the nonlabeled features. In the particular case of Experiment 1, labeling the object identity seems to have led to the filtering of the color information.

In contrast to a labeling benefit in visual working memory, we could not find evidence for a labeling benefit for the retention of the same objects for a delayed recall (visual long-term memory) test in Experiments 1 and 2. This was the case when participants were not aware (Experiment 1a) and aware of the visual long-term memory test (Experiment 1b) prior to the start of the experiment. Moreover, in Experiment 2, we ruled out the possibility that this lack of effect was due to rather poor visual long-term memory in general. In Experiment 2, participants repeatedly saw the same color-object pairs for three consecutive trials, thereby fostering learning in visual long-term memory by means of the repetition (Couture & Tremblay, 2006; Johnson et al., 2017; Lafond et al., 2010; Oberauer & Meyer, 2009) and the testing-effect (Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger & Pyc, 2012; Sutterer & Awh, 2016). This manipulation substantially improved delayed recall, yet no labeling benefit was observed for visual long-term memory. These findings are in line with the lack of a color labeling benefit observed by Kelly and Heit (2017).

These experiments suggest that the beneficial effect of labeling on visual working memory observed by Souza and Skóra (2017) and the lack of a labeling benefit on visual long-term memory observed by Kelly and Heit (2017) are not attributable to differences in the procedures used to induce labeling (aloud responses vs. keypress) and to test memory (continuous color reproduction vs. color-hue recognition test). Here, we maintained these features constant and were able to show the same dissociation in the retention of labeled information over the short and long term. The only experiment in which we could obtain a labeling benefit both in visual working memory and visual long-term memory was Experiment 3, wherein the color and object identity were labeled together.

Implication of Verbal Labeling for Continuous and Categorical Representations in Memory

One aim of the present study was to analyze the contribution of verbal labeling to the storage of coarse (categorical) and more fine-grained (continuous) visual representations over the short and long term. This was assessed by modeling the data with a mixture model that attempts to distinguish between the sources of information used to respond in the task, namely categorical information about the colors, continuous information about the precise hue studied (and the precision of this information) or guessing.

For visual working memory, mixture modeling of all experiments indicated that labeling the color of an object increased the probability of retrieval of this information overall as opposed to guessing, replicating Souza and Skóra (2017). These authors further showed that this benefit was not solely attributable to addition of categorical representations: either the probability of continuous information in memory increased while continuous precision remained relatively the same; or continuous precision increased along with little change in the amount of continuous information stored. In the present study, we found a similar mix of effects: the quantity of continuous memory increased in Experiment 3, whereas we found rather improvements in memory precision in Experiments 1 and 2. This means that labeling allowed detailed information from a larger number of items to be stored (e.g., effect on continuous memory parameter), or that the number of items for which continuous information was retained remained the same, but their continuous recall was more precise (e.g., effect on continuous imprecision parameter).

To the best of our knowledge, the categorical-continuous mixture modeling approach has not yet been used to assess visual long-term memory nor the role of labeling therein. Our experiments showed that, in general, information stored in visual long-term memory had a lower probability of retrieval and lower precision compared with visual working memory, replicating prior findings (Biderman et al., 2019). The lower visual long-term memory precision was observed although the model controls for categorical responding, which in itself would be associated with lower precision in mixture models that do not include categorical responses. This shows that the lower precision of visual long-term memory representations cannot be accounted by larger proportion of categorical responses in delayed tests. Furthermore, in Experiments 1a, 1b and 3, we also observed that the probability of retrieving categorical representations was higher than of retrieving continuous representations in visual long-term memory, whereas for visual working memory the division between categorical and continuous representations was more even. This suggests that another differentiating factor between visual long-term memory and visual working memory may pertain to the retention of continuous information. It is also worth noting that estimates of continuous representations in visual long-term memory were generally low (ca. 5%, ranging between 4 and 8%) across Experiments 1a, 1b, and 3. In these experiments, no repetitions were implemented, and participants learned a large set of items, namely 315 colored objects. This is consistent with an average of 14 objects retrieved with continuous information. In contrast, estimates of categorical representations were two to four times larger (ranging from 10% to 32%), indicating that participants could retrieve many more coarse representations in visual long-term memory. The lower fidelity of visual long-term memory, however, seems related to the limited opportunities to commit visual information to this system. Experiment 2 showed that repetitions improved delayed recall, substantially increasing the probability of continuous information storage and its precision. These results corroborate the findings of Miner et al. (2020), indicating that representations in long-term memory can also have high fidelity provided that multiple traces of the object have been stored.

Regarding labeling, no effect was observed on visual long-term memory when the data was modeled in Experiments 1a and 1b in agreement with the results obtained for the model-free index of performance (i.e., recall error). In Experiment 2, labeling tended to increase categorical information at the expense of more continuous information (a small and noncredible reduction on probability of retrieval and on precision), such that average performance did not improve (as indicated by the recall error data). In Experiment 3, labeling improved performance as revealed by the recall error analysis, but again mixture modeling indicated that this benefit was associated with increases in categorical memory only (with continuous memory remaining unchanged), unlike what was observed for visual working memory. Hence, labeling of both features seems to play an important role for a labeling benefit that can be maintained across a longer time-period into visual long-term memory. The novel insight provided by this experiment was that the labeling benefit in visual long-term memory reflected an increase in categorical information with no change in continuous memory, again in stark contrast to the effects observed for visual working memory.

Different Role of Labels in Visual Working Memory and Visual Long-Term Memory

In the introduction we discussed five hypotheses of the effect of verbal labeling in visual memory. Our results help distinguishing between the plausibility of these hypotheses as likely explanations of the labeling effect in visual working memory and visual long-term memory.

First, our results do not support the *verbal recording* hypothesis, neither for visual working memory nor visual long-term memory: Across most experiments, we did not find an indication that labeling increased categorical representations at the expense of continuous information or its precision as predicted by this hypothesis. For visual long-term memory, some studies have found a cost for labeling in line with the verbal overshadowing effect or memory distortion effect (Brandimonte et al., 1997; Lupyan, 2008; Schooler & Engstler-Schooler, 1990). The only instance in which we observed a tendency for a trade-off between categorical and continuous information in visual long-term memory was in Experiment 2. This trend was not credible though.

Second, the *dual-trace* hypothesis predicts that labeling would only increase categorical responding with no change in continuous memory. This prediction fits with the labeling benefit observed for visual long-term memory in Experiment 3. This hypothesis, however, cannot explain the visual working memory data.

Third, the *distinctiveness* hypothesis predicts that the labeling benefits would be proportional to how much the label differentiates between the memoranda. In Experiments 1a and 1b, we included an object labeling condition that allowed the generation of a unique label for each item in the experiment (since each object was only presented once) which adds more distinctiveness to the memory traces than the color labeling condition. Contradicting this hypothesis, the object-labeling condition yielded costs to visual working memory performance, and no effect for visual long-term memory retrieval.

Fourth, the *activation of categorical visual long-term memory* hypothesis predicts that labels activate visual long-term memory representations of the category. This would allow people to store more visual details because the individual item's properties can be stored in relation to the category. This may facilitate data compression or the use of hierarchical representations that reduce memory load (Brady et al., 2009). In line with this hypothesis, visual working memory performance benefited from color labeling by showing an increase in continuous memory or continuous precision (see also Forsberg et al., 2020; Souza & Skóra, 2017). This effect, however, was constrained to visual working memory; visual long-term memory did not show increases in continuous memory as a function of labeling.

Fifth, the *cue to focus attention* hypothesis (Kelly & Heit, 2017) predicts that labeling guides attention to the labeled features, and this can be helpful or harmful depending on the match between the attended and the relevant feature. In our experiments, participants were fully aware that color was the relevant feature, thus color labeling could not be beneficial according to this hypothesis. Object labeling, however, would direct attention away from the relevant feature and hence this hypothesis predicted a cost in this condition. Our data partially matches those predictions: On the one hand, this hypothesis fails to account for the fact that color labeling does improve memory, especially visual working memory, but also visual long-term memory if color labeling is combined with labeling the object. On the other hand, it correctly predicts a cost for object labeling in visual working memory. This suggests that labels serve to guide attention to certain features, but this does not fully explain the resulting benefits that follow from it. The hypothesis as formulated by Kelly and Heit (2017), however, disregards the possibility that labeling may increase the amount of attention toward the labeled information or the amount of time attention dwells on it, thereby increasing memory performance when the labeled feature is the relevant feature even when participants are fully aware of it. A reformulation of this hypothesis along these lines could account for our data. Future studies are therefore needed to assess how much attention is engaged during labeling and whether nonlabeling conditions matched on attention engagement could yield the same benefits we observed here.

To conclude, we found evidence in partial support of three mechanisms: (a) verbal labels guide attention to the labeled feature, and this differential attention affects visual working memory processing, (b) the label activates categorical knowledge in visual long-term memory, and (c) for visual working memory, this visual long-term memory activation allows for storage of more visual details, perhaps because categorical information permits exploitation of redundancies in the visual input (e.g., facilitating data compression or creation of hierarchical representations) or reducing interitem interference. These more precise representations created in visual working memory, however, either are (a) not transferred to visual long-term memory or (b) they do not seem to survive the proactive interference accumulated in visual long-term memory as more and more objects are learned. As such, at best, knowledge activation through labels only serves to increase categorical storage in visual long-term memory, and only if this activation is combined with the concomitant activation of the retrieval cue (e.g., the object's label).

Creation of Representations in Visual Working Memory and Visual Long-Term Memory

To create a durable memory representation, the visual object needs to be perceived, encoded, and consolidated to be later accessible in memory (Cowan, 2017; Ricker, 2015). Attention and time are assumed to be necessary to create stable memory representations both in visual working memory (Ricker & Cowan, 2014; Ricker & Hardman, 2017) and visual long-term memory (Huebner & Gegenfurtner, 2011). During encoding, a visual trace of the memory object is built up, which is then transformed into a memory representation by the process of consolidation (Ricker, 2015; Ricker, Nieuwenstein, Bayliss, & Barrouillet, 2018). So far, it is unclear whether consolidation creates a representation that is accessible both over the short and the long term or whether there are separate consolidation processes operating in each memory system. Facilitation of consolidation in visual working memory could explain the short-term benefits of labeling. Consolidation is known to continue even after the offset of the memory item (Ricker & Hardman, 2017), which constitutes the critical period in which labeling occurred in our experiments. Labeling may have facilitated the creation of a stable representation in visual working memory of the continuous and categorical information available in the sensory stimulus. If labeling improves memory by facilitating short-term consolidation, this would suggest that consolidation in long-term memory is likely a separate process. Our findings, therefore, are relevant to the understanding of the interplay of working memory and long-term memory.

Memory models make different assumptions about the relation between working memory and long-term memory. One line of models assumes that working memory (termed short-term memory [STM] at the time) and long-term memory represent distinct stores with bidirectional interactions. Representations of external inputs first enter working memory, and only after that can be transferred to long-term memory (Atkinson & Shiffrin, 1971). Information in long-term memory can also be activated and then transferred to working memory to facilitate processing, as is the case when prior knowledge facilitates immediate memory. Critically, this view conceptualizes working memory as the gateway to the creation of representations in long-term memory. For Atkinson and Shiffrin (1971), time in working memory was determinant for successful transfer of information to long-term memory. Others have proposed that the depth of the processing was the factor that established information in the long-term memory store (Craik & Lockhart, 1972). Our findings place some challenges to these theories. First, we observed that improving or hindering recall from working memory had virtually no effect on how likely or precisely information was retrieved from long-term memory. This challenges the gateway hypothesis. Second, labeling can be viewed as a task that increases the depth of processing of the stimuli compared with saying irrelevant syllabi aloud (as in the revised Cue to Attention hypothesis), yet it did not improve memory over the long term.

Other sets of models conceptualize working memory as consisting of multiple components (i.e., the multiple component model; Baddeley, 1986, 2012, 2017; Baddeley & Logie, 1999; Logie, 2011). These models assume that visual and verbal information are stored in separate buffers, with their own rehearsal mechanisms (Baddeley, 2012; Logie, 2011). These models predict that information stored in more than one code (e.g., both as visual and verbal traces) would have increased chance of recall because of the added capacities of using two separate buffers (Logie, 2018; Logie et al., 2016), in line with the dual-coding model (Paivio, 1971). Accordingly, these models predict a benefit of labeling in terms of increases in categorical representations but have difficulty explaining how labeling impacts storage of continuous information. One way in which the multicomponent model could deal with this interaction is via the assumption of an episodic buffer that combines representations from different working memory stores and from long-term memory (Baddeley, 2000, 2012). How exactly information is combined from these different modalities and how it is integrated with long-term memory is underspecified in the model. As in the previously described models, working memory is usually viewed as a starting point for establishing information in long-term memory, with this contribution being larger for novel information (e.g., for learning of nonwords; Baddeley et al., 1988). Because the model is silent about how exactly the flow of information between working memory components and long-term memory occurs in the episodic buffer, it is unclear how it would account for the differential effect of labeling over the short and long term we observed.

Other models view working memory as an activated subset of long-term memory representations, with its capacity limitations arising in maintaining relevant information within a broad focus of attention that keeps accessible a small number of chunks (Cowan, 1988; Oberauer, 2009; Oberauer & Hein, 2012). Oberauer and Hein (2012) proposed that the broad focus of attention could be further divided into a narrow focus holding only one single object, thereby giving it a special role. Within such models, verbal labeling can be conceived as a further way to activate representations in visual long-term memory (besides the activation induced by the visual input itself), thereby facilitating the binding of information to their relevant context in the broad and narrow focus of attention. Although activation in long-term memory is usually assumed to be unlimited and to spread to nearby nodes (Oberauer, 2009), it is conceivable that activation of categories via labeling might boost the most relevant feature values, reducing interference from previously encoded stimuli, and facilitating the creation/consolidation of a binding between the precise stimulus and its context. Why does this categorical activation not facilitate long-term learning? One explanation might rely on the spread of activation within this system and the build-up of interference. Because so many representations are stored in long-term memory, retrieval from this system is based on a slow and error-prone cue-based search. Having larger activation of color categories is not sufficient to facilitate search through the hundreds of objects stored in long-term memory, particularly if this activation was not enough to establish a robust binding between the object and its color. This may explain why labeling the object and the color was necessary to observe some long-term learning: this permitted the storage of stronger color-object bindings that could be used to search through memory after a delay, although only categorical information survived the build-up of interference from encoding multiple objects.

Verbal Labeling Benefit in Relation to Retrieval Practice

In Experiment 2 we replicated the beneficial effect of repeated studying and testing on both visual working memory and visual long-term memory (Roediger & Butler, 2011; Roediger & Karpicke, 2006; Roediger & Pyc, 2012; Sutterer & Awh, 2016). In the visual working memory task, each color-object pair was presented and tested three times while participants labeled the colors or said “bababa” aloud throughout the repetitions. The labeling benefit was restricted to the very first exposure to the object and vanished for the second and third repetition, contributing further evidence that the verbal labeling effect is short-lived and does not affect the rate of learning.

Relatedly, the absence of a verbal labeling effect for visual long-term memory in general, and with the repeated presentation of the colored objects rules out an explanation of the verbal labeling effect as retrieval practice. One could argue that to label, one has to retrieve this information, thereby leading to an additional retrieval practice not present in the suppression condition. This retrieval practice could explain the beneficial effects of labeling observed by Souza and Skóra (2017) and in Experiment 1. If this was the case, we should expect labeling to improve visual long-term memory since we know retrieval practice does improve visual long-term memory retention (Sutterer & Awh, 2016). Furthermore, performance in the second presentation of the colored object in the labeling condition would imply four retrievals (two in the first trial + two in the second trial), and hence it should have been even better than performance in the third repetition in the suppression condition. Experiment 2 showed, however, that performance improved linearly with the number of repetitions in visual working memory irrespective of labeling (see also Miner et al., 2020). This is inconsistent with the possibility that labeling benefits visual working memory through retrieval practice.

Conclusion

The way in which we describe our visual surroundings can have a profound impact on the visual memories that are formed to guide our behavior over the short and long term. Here we demonstrated for the first time that verbal labeling is either beneficial or inconsequential for the retention of visual memories and that the source of this benefit is different across short and long timescales. Verbal labels provide categorical information that boosts the maintenance of high-fidelity representations in visual working memory to guide our immediate behavior. These detailed representations are either not retained over the long term or they do not survive interference that accumulates in visual long-term memory. As such, verbal labeling can, at best, allow for the retention of more categorical information over the long term.

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š, Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., Brown, C., Buswell, K., Carlson, C., Carlson, M., Chu, S., Cislak, A., Colarusso, M., Colloff, M. F., Dellapaolera, K. S., Delvenne, J.-F., . . . Zwaan, R. A., (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578. https://doi.org/10.1177/1745691614545653

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *225*(2), 82–90. https://doi.org/10.1038/scientificamerican0871-82

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

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. (Ed.). (2017). The concept of working memory: A view of its current state and probable future development. *Exploring working memory* (pp. 99–106). Routledge. https://doi.org/10.4324/9781315111261-9

Baddeley, A., & Logie, R. H. (1999). *Working memory: The multiple-component model*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*(5), 586–595. https://doi.org/10.1016/0749-596X(88)90028-9

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. https://doi.org/10.1037/xge0000076

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working memory has better fidelity than long-term memory: The fidelity constraint is not a general property of memory after all. *Psychological Science*, *30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Blanco, N., & Gureckis, T. (2013). Does category labeling lead to forgetting? *Cognitive Processing*, *14*(1), 73–79. https://doi.org/10.1007/s10339-012-0530-4

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

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*(5), 4, 4. https://doi.org/10.1167/11.5.4

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981–990. https://doi.org/10.1177/0956797612465439

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, *20*(4), 449–455. https://doi.org/10.3758/BF03210929

Brandimonte, M. A., Schooler, J. W., & Gabbino, P. (1997). Attenuating verbal overshadowing through color retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 915–931. https://doi.org/10.1037/0278-7393.23.4.915

Brown, C., Brandimonte, M. A., Wickham, L. H., Bosco, A., & Schooler, J. W. (2014). When do words hurt? A multiprocess view of the effects of verbalization on visual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1244–1256. https://doi.org/10.1037/a0037222

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, *15*(1), 73–86. https://doi.org/10.1037/h0072671

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory & Cognition*, *34*(8), 1720–1729. https://doi.org/10.3758/BF03195933

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178. https://doi.org/10.3758/s13423-014-0675-5

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123. https://doi.org/10.1037/xge0000560

Forsberg, A., Johnson, W., & Logie, R. H. (2020). Cognitive aging and verbal labeling in continuous visual memory. *Memory & Cognition*, *48*(7), 1196–1213. https://doi.org/10.3758/s13421-020-01043-3

Hardman, K. O. (2016). CatContModel: Categorical and continuous working memory models for delayed estimation tasks (Version 0.7.5) [Computer Software]. https://github.com/hardmanko/CatContModel/releases/tag/v0.7.5

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*(1), 30–54. https://doi.org/10.1037/xhp0000290

Huebner, G. M., & Gegenfurtner, K. R. (2011). The efficiency of encoding: Limits of information transfer into memory *Attention, Perception & Psychophysics*, *73*(5), 1503–1521. https://doi.org/10.3758/s13414-011-0120-z

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. *Memory*, *25*(9), 1279–1293. https://doi.org/10.1080/09658211.2017.1293692

Kelly, L. J., & Heit, E. (2017). Recognition memory for hue: Prototypical bias and the role of labeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 955–971. https://doi.org/10.1037/xlm0000357

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Lafond, D., Tremblay, S., & Parmentier, F. (2010). The ubiquitous nature of the Hebb repetition effect: Error learning mistaken for the absence of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 515–522. https://doi.org/10.1037/a0018469

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*, *20*(4), 240–245. https://doi.org/10.1177/0963721411415340

Logie, R. H. (2018). Human cognition: Common principles and individual variation. *Journal of Applied Research in Memory & Cognition*, *7*(4), 471–486. https://doi.org/10.1016/j.jarmac.2018.08.001

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition*, *44*(4), 590–607. https://doi.org/10.3758/s13421-015-0580-9

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369. https://doi.org/10.1037/0096-3445.137.2.348

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

MacLeod, C. M. (2010). When learning met memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *64*(4), 227–240. https://doi.org/10.1037/a0021699

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Miner, A. E., Schurgin, M. W., & Brady, T. F. (2020). Is working memory inherently more "precise" than long-term memory? Extremely high fidelity visual long-term memories for frequently encountered objects. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(8), 813–830. https://doi.org/10.1037/xhp0000748

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.12-2) [Computer software]. https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, *21*(3), 164–169. https://doi.org/10.1177/0963721412444727

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory*, *17*(7), 774–781. https://doi.org/10.1080/09658210903107861

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Ozubko, J. D., & Macleod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1543–1547. https://doi.org/10.1037/a0020604

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, *88*, 1–21. https://doi.org/10.1016/j.cogpsych.2016.05.003

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 261–282. https://doi.org/10.1037/0096-1523.24.1.261

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1579–1587. https://doi.org/10.1037/a0024347

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language*, *68*(1), 10–25. https://doi.org/10.1016/j.jml.2012.09.001

Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *Neuroscience*, *2*(4), 259–279. https://doi.org/10.3934/Neuroscience.2015.4.259

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 417–428. https://doi.org/10.1037/a0034301

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*, *146*(11), 1551–1573. https://doi.org/10.1037/xge0000346

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: Insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 8–18. https://doi.org/10.1111/nyas.13633

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory & Cognition*, *1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304–321. https://doi.org/10.1037/met0000057

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, *49*(3), 853–862. https://doi.org/10.3758/s13428-016-0741-1

Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Journal of Memory and Language*, *119*, 104242. https://doi.org/10.1016/j.jml.2021.104242

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin & Review*, *23*(3), 831–841. https://doi.org/10.3758/s13423-015-0937-x

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics, always learning*. Pearson.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. https://doi.org/10.1167/4.12.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

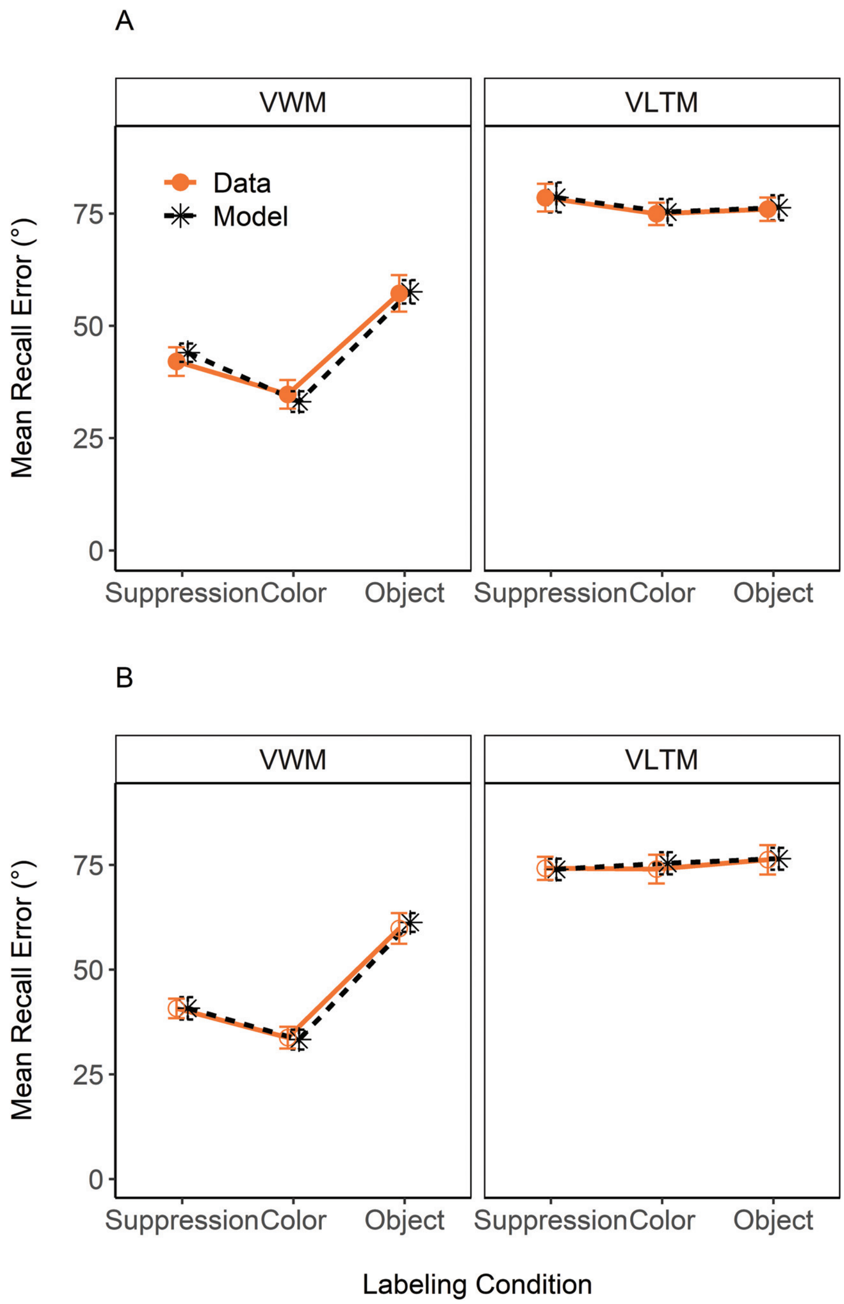
Appendix

Model Fit

To assess how well the model captured the data, a posterior predictive check was performed by simulating data (predictions) based on the full model parameters for all experiments. Figure A1A and A1B shows that the predicted recall error seemed to be fairly in line with the data for Experiment 1a and 1b, respectively. For Experiment 2, Figure A2A shows that the modeling fit the data for Experiment 2 for the three repetitions in visual working

(*Appendices continue*)

Figure A1
Recall Error Obtained for the Data of Experiments 1a and 1b and the Predicted, Simulated Data From the Posterior Estimates of the Mixture Model Fitted to These Data

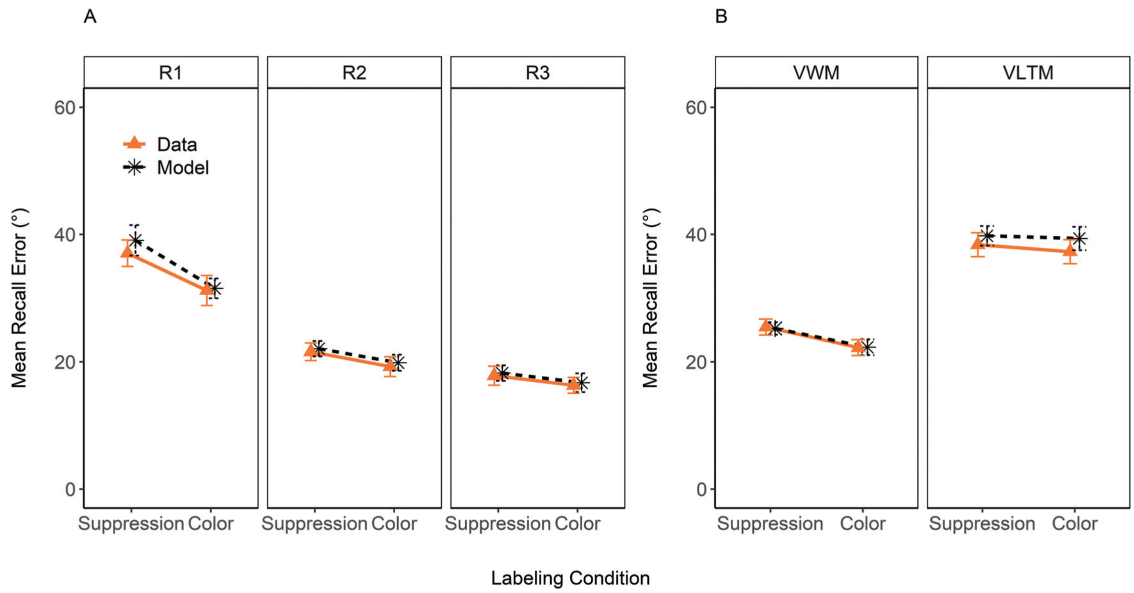


Note. Error bars represent the 95% within-subjects confidence interval. VWM = visual working memory; VLTM = visual long-term memory. See the online article for the color version of this figure.

(*Appendices continue*)

Figure A2

Recall Error Obtained for the Data of Experiment 2 and the Predicted, Simulated Data From the Posterior Estimates of the Mixture Model Fitted to These Data

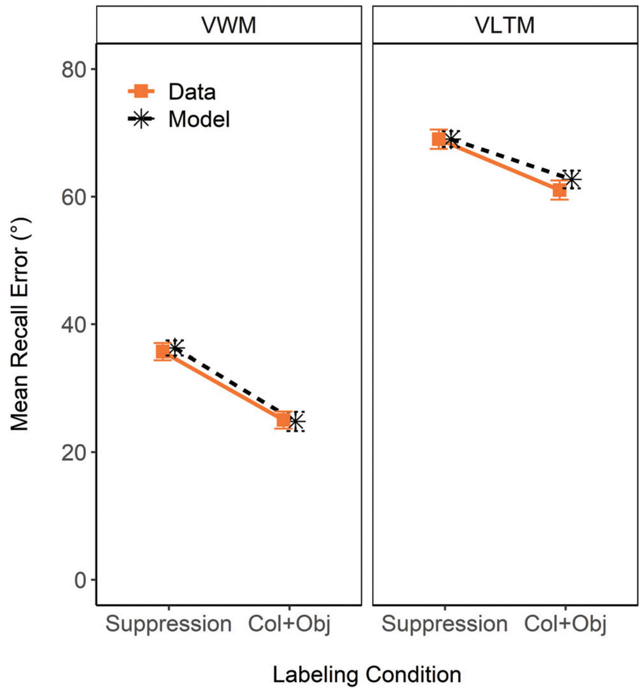


Note. Panel A shows the data and predictions for the trial repetitions (R1, R2, R3) as a function of labeling condition in the visual working memory phase. Panel B shows data and predictions for the model comparing visual working memory performance (averaged across repetitions) and visual long-term memory. Error bars represent 95% confidence interval. VWM = visual working memory; VLTM = visual long-term memory. See the online article for the color version of this figure.

(*Appendices continue*)

Figure A3

Recall Error Obtained for the Data of Experiment 3 and the Predicted, Simulated Data From the Posteriors of the Mixture Model Fitted to These Data



Note. Error bars represent 95% within-subjects confidence intervals. See the online article for the color version of this figure.

memory, and Figure A2B for the visual working memory and visual long-term memory model. Figure A3 shows that the posterior estimates of the model in Experiment 3 also reproduced the actual data.

Received June 15, 2020
Revision received February 26, 2021
Accepted April 5, 2021 ■