Cognition 166 (2017) 277–297

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original articles

The interplay of language and visual perception in working memory ☆

Alessandra S. Souza [a,*], Zuzanna Skóra [b]

[a] Department of Psychology, University of Zurich, Switzerland
[b] Institute of Psychology, Jagiellonian University, Poland

ARTICLE INFO

Article history:
Received 1 June 2016
Revised 17 May 2017
Accepted 30 May 2017

Keywords:
Categorical memory
Continuous features
Language
Visual working memory

ABSTRACT

How do perception and language interact to form the representations that guide our thoughts and actions over the short-term? Here, we provide a first examination of this question by investigating the role of verbal labels in a continuous visual working memory (WM) task. Across four experiments, participants retained in memory the continuous color of a set of dots which were presented sequentially (Experiments 1–3) or simultaneously (Experiment 4). At test, they reproduced the colors of all dots using a color wheel. During stimulus presentation participants were required to either label the colors (color labeling) or to repeat "bababa" aloud (articulatory suppression), hence prompting or preventing verbal labeling, respectively. We tested four competing hypotheses of the labeling effect: (1) labeling generates a verbal representation that overshadows the visual representation; (2) labeling yields a verbal representation in addition to the visual one; (3) the labels function as a retrieval cue, adding distinctiveness to items in memory; and (4) labels activate visual categorical representations in long-term memory. Collectively, our experiments show that labeling does not overshadow the visual input; it augments it. Mixture modeling showed that labeling increased the quantity and quality of information in WM. Our findings are consistent with the hypothesis that labeling activates visual long-term categorical representations which help in reducing the noise in the internal representations of the visual stimuli in WM.

1. Introduction

The present study is concerned with how visual perception and language interact to form the representations that guide our thoughts and actions over the short-term. The memory system holding information accessible for the moment-to-moment cognition is known as working memory (WM). In WM research, the mainstream strategy has been to study processing of visual and verbal inputs in isolation. In contrast to the laboratory, day-to-day observations suggest a more interactive scenario in which visual inputs and language co-exist and interact. For example, in order to safely change lanes, one has to locate the positions of the other cars, check for traffic signs, and look for potential pedestrians. In each of these steps, one may generate or receive verbal descriptions of the ongoing events. How are these incoming inputs combined in mind to effectively guide action? At the moment, we lack a systematic treatment of the consequences of having both visual and verbal inputs regarding the same event to guide behavior over the short-term. Here, we provide a first examination of this question by investigating the role of verbal labeling for the temporary retention and retrieval of visual inputs varying along a continuous dimension.

The retention of continuous feature values in memory can be studied with the *continuous delayed estimation task* (Prinzmetal, Amiri, Allen, & Edwards, 1998; Wilken & Ma, 2004; Zhang & Luck, 2008). Color reproduction has received the largest attention in the visual WM literature (Allred & Flombaum, 2014). In a typical WM color delayed-estimation task, participants have to retain the precise color-hues of an array of objects. At test, the hue of a target object has to be reproduced using a continuous color wheel. The dependent measure in this task is recall error computed as the distance between the reported value and the target's true value. The more precise the representation of the studied items in memory, the smaller the error in reproducing the target's feature. Furthermore, the distribution of responses in this task can be submitted to mixture modeling to estimate the probability that responses were informed by memory as opposed to guessing, and the variability (imprecision) with which this information was stored (cf. Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008). The sensitivity of this task to the quality of the underlying visual representation

☆ This research was supported by a grant from the Forschungskredit (FK-13-083) and from the Swiss National Science Foundation (project no. 169302) to Alessandra S. Souza.

* Corresponding author at: Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland.
E-mail address: a.souza@psychologie.uzh.ch (A.S. Souza).

http://dx.doi.org/10.1016/j.cognition.2017.05.038
0010-0277/

makes it a perfect testbed to assess changes in visual WM as a function of verbal labeling.

In standard visual WM tasks, all of the visual stimuli are presented in a one-shot display for a very brief interval (a few hundred milliseconds), and memory is tested shortly after (typically 1 s). The fast pace with which the trial progresses, and the larger number of items displayed simultaneously strongly discourages verbal labeling. This is corroborated by the finding that in change detection tasks (which require the recognition of one of the displayed items) further blocking labeling with the addition of a verbal memory load (cf. Vogel, Woodman, & Luck, 2001) or by asking participants to articulate irrelevant words continuously (aka. *articulatory suppression*) has no impact on performance (Morey & Cowan, 2004, 2005; Sense, Morey, Prince, Heathcote, & Morey, 2016).

To the best of our knowledge, only one study considered how labels affect performance in a continuous delayed estimation task. Donkin, Nosofsky, Gold, and Shiffrin (2015) asked participants to store the precise color of a single dot presented for 0.1, 0.5, or 2 s. In color reproduction trials, stimulus offset was followed by a varying retention interval, after which memory was tested with a color wheel. In labeling trials, following stimulus offset, participants were asked to type a label to the color. Three trials later, the label was presented onscreen together with the color wheel, and participants had to pick the color represented by the label. Longer study durations yielded more precise perceptual memory of the stimulus, as well as more precise responding in labeling trials. Modeling of responses in color reproduction trials showed that a mixture of perceptual information, verbal labeling, and of random guessing best fitted the data. Moreover, the model incorporating decay of the visual input over the retention interval, with no decay of the verbal representation, also fitted best. In the study of Donkin et al., the precision of labeling responses was directly assessed in labeling trials, and including this information in modeling improved fitting. Still, this study provided no means to distinguish for the differential impact of labeling because no condition was included that prevented labeling from occurring.

In sum, the extant visual WM literature does not inform us about the possible consequences of allowing people to generate descriptions of the visual events they experience. To tackle this question, we developed an overt labeling protocol to strictly control the labeling opportunities for each item in the memory display. In our experiments, we presented items sequentially, and provided sufficient time after each item to allow for labeling. In the labeling condition, participants were prompted to label each presented item aloud. In the control condition, participants were required to constantly articulate "*bababa*" aloud (articulatory suppression). The articulatory suppression procedure prevents people from articulating and hence generating verbal labels.

Armed with a proper task set-up, we conducted four experiments in which we manipulated further variables to illuminate the space of explanations of the interplay between verbal and visual inputs in memory. In the following, we will delineate four hypotheses of the labeling effect in visual WM that we aimed at distinguishing empirically. We will substantiate the plausibility of these hypotheses with findings from the effects of language on several aspects of cognition, from perception to episodic memory.

1.1. Verbal recoding

The first possibility is that labeling generates a verbal representation at the expense of the visual one. Storage of the term "*green*" at the expense of the particular greenish hue presented for study should lead to a large loss of precision in recalling this feature from WM. There is evidence that verbalizations can hamper visual long-term memory (LTM). One piece of evidence comes from the verbal overshadowing effect, namely the observation of worse memory for a face (or even color) when in between study and test participants are asked to describe the stimulus (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990). In a related vein, it has been observed that asking people to classify objects as being one out of 2 categories (e.g., lamps vs. chairs) impairs LTM for the studied exemplars, compared to asking for preference ratings (Lupyan, 2008). These studies suggest that verbal descriptions can hamper visual LTM, possibly due to the loss of the visual trace. It is unclear whether a similar effect is observed in visual WM.

1.2. Dual-trace (visual + verbal)

The second possibility is that labeling adds a verbal representation to the visual trace in WM. It follows that participants would have two sources of information: a continuous visual representation, and a verbal label. The joint information from both traces could be combined during recall to yield the final response output; or one of the two representations may dominate depending on the test situation. A dual-trace hypothesis guided the modeling in the study by Donkin et al. (2015): by entering verbal labeling as an additional source of information in mixture modeling, the authors assumed that both visual and verbal inputs co-existed in WM and interacted in guiding reproduction of colors from memory.

Support for this dual-trace hypothesis comes from studies of LTM memory for easy-to-label and hard-to-label drawings performed by Brandimonte and colleagues. They showed that labeling can hamper visual LTM memory, consistent with a verbal overshadowing effect (Brandimonte, Hitch, & Bishop, 1992); however, this effect can be reversed if the visual context for the studied item (e.g. its color) is reinstated at test (Brandimonte, Schooler, & Gabbino, 1997), and the impairing effect depends on the match between the type of verbal description (one label vs. description of features) and the information required at test (global or feature-based) (Brown, Brandimonte, Wickham, Bosco, & Schooler, 2014). They have also shown that verbal descriptions may be beneficial when generated in the presence of the stimulus, but not during a retention interval (Nakabayashi, Mike, Brandimonte, & Lloyd-Jones, 2012). These findings suggest that labeling may yield a verbal representation in memory in addition to the visual input, and that stronger reliance on either type of representation can be varied depending on the retrieval cues presented at test.

1.3. Distinctiveness

The third possibility is that generating a label benefits visual WM because it yields an additional retrieval cue to the labeled item. If participants associate labels with the visual representations, and they remember the pairing of the labels to the spatial locations of items at test, the label can be used to more effectively retrieve the continuous visual representations from WM. If this holds, it would indicate that labeling helps to the extent that it distinguishes between items in memory. According to this account, the LTM impairment observed by Lupyan (2008) could be explained by the lack of distinctiveness of the labels used in this study (only 2 labels for several exemplars of the same category). Richler, Palmeri, and Gauthier (2013) found that LTM for vocally labeled objects (from different categories) was similar to memory for items for which participants made preference ratings, and both conditions yielded better LTM than silent study of the objects. Furthermore, labeling improved rejection of both within-category and between-category lures, hence indicating better memory for the specific exemplars studied. In addition, generating the labels aloud (as opposed to typing them) also played a role, implying a contribution of mode of production to this effect.