Acta Psychologica 167 (2016) 45–51

Contents lists available at ScienceDirect

Acta Psychologica

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

To bind or not to bind, that's the wrong question: Features and objects coexist in visual short-term memory

Shriradha Geigerman [a], Paul Verhaeghen [a], John Cerella [b]

[a] Georgia Institute of Technology, United States
[b] Syracuse University, United States

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form 11 April 2016
Accepted 13 April 2016
Available online 22 April 2016

Keywords:
Working memory
Visual short-term memory
Binding
Objects
Features

ABSTRACT

In three experiments, we investigated whether features and whole-objects can be represented simultaneously in visual short-term memory (VSTM). Participants were presented with a memory set of colored shapes; we probed either for the constituent features or for the whole object, and analyzed retrieval dynamics (cumulative response time distributions). In our first experiment, we used whole-object probes that recombined features from the memory display; we found that subjects' data conformed to a kitchen-line model, showing that they used whole-object representations for the matching process. In the second experiment, we encouraged independent-feature representations by using probes that used features not present in the memory display; subjects' data conformed to the race-model inequality, showing that they used independent-feature representations for the matching process. In a final experiment, we used both types of probes; subjects now used both types of representations, depending on the nature of the probe. Combined, our three experiments suggest that both feature and whole-object representations can coexist in VSTM.

1. Introduction

Visual short-term memory (VSTM) stores visual information for a brief period of time (a few seconds) in the interest of ongoing tasks (Baddeley & Hitch, 1974). There is considerable consensus that a severe capacity limit exists on the number of items it can represent – typically no more than 3 or 4 at a time (Cowan, 2001; Luck & Vogel, 1997).

A question central to our understanding of the observed capacity limit within VSTM is qualitative: What is the format of representation of information in VSTM (e.g., Suchow, Fougnie, Brady, & Alvarez, 2014) that determines its capacity limits? There are three points of view on this matter. According to the *object-driven view*, VSTM stores information in the form of integrated objects; capacity is then measured as the number of (multi-feature) objects that can be maintained simultaneously (Lee & Chun, 2001; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). Alternatively, the *feature-driven view* of VSTM posits that capacity is influenced by the total number of features (such as colors, shapes, or orientations) that constitute a visual scene (Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Wheeler & Treisman, 2002), independent of the number of objects across which those features are distributed. The third view is a hybrid: The *object-and-features-driven view* proposes dual representation of both objects and features in VSTM. According to this view, the VSTM capacity limit is a function of both the number of objects and the number of features that can be accurately maintained at a given time (e.g. Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013; Vergauwe & Cowan, 2015; Wheeler & Treisman, 2002).

Support for the object-driven view rests on findings from the change detection paradigm. In this paradigm, participants are presented with two consecutive displays separated by either a blank display or a mask (colored shapes are a favorite type of stimulus). The two displays are either identical, or they differ in one item; participants indicate whether or not a change has occurred. The typical result is that accuracy is high for 3 to 4 items, but declines sharply at larger set sizes (e.g. Cowan, 2001; Luck & Vogel, 1997; Vogel et al., 2001; Wheeler & Treisman, 2002). Research that supports the feature-driven view, on the other hand, typically uses a reconstruction paradigm. In this paradigm, features of a more continuous nature are used (e.g., shades of green, line orientations, Gabor patches, and the like), and participants have to reconstruct one or more items from the memory array by 'dialing in' their response (Bays & Husain, 2008; Bays et al., 2011; Wilken & Ma, 2004). Finally, objects-and-features studies tend to use the change-detection paradigm, now manipulating feature characteristics, such as complexity (Alvarez & Cavanagh, 2004) or the number of within-object features (Oberauer & Eichenberger, 2013) in addition to the number of objects.

Some recent research has made it increasingly clear that the strict object-driven and feature-driven views are too restrictive. Notably, at least four papers (all using the change detection paradigm) have provided evidence for the position that people can selectively encode, store, and/or retrieve either features or objects in VSTM, depending on the requirements of the task. These studies thus directly refute the notion that memory representations consist only of objects, or only of unbound features. The first study to demonstrate selectivity was by

http://dx.doi.org/10.1016/j.actpsy.2016.04.004
0001-6918/

Woodman and Vogel (2008). In their experiment, participants were shown colored bars in different orientations (their Expt. 1) or colored objects of different shapes (their Expt. 2); they were instructed to remember either color or orientation or both (their Expt. 1), or either color or shape or both (their Expt. 2). The main result was that memory was better for the color-feature alone condition than for the conditions where color was combined with either orientation or shape. This strongly suggests that when participants are instructed to hold on to only a single feature, they can do so. Kondo and Saiki (2012) obtained a similar result. They asked their subjects to pay attention to the conjunction of two of three features of stimuli varying in color, shape, and location (color-shape, shape-location, or color-location combinations). Subjects were able to block interference from the task-irrelevant feature (only, however, if it was shape or color, not if it was location), suggesting that they were able to selectively filter out some types of information (viz., color or shape) as required. Morey, Guérard, and Tremblay (2013) compared memory for either the color or the letter identity of colored letters under three conditions: a reinstated stimulus (e.g., if the display contained a red 'M' and a blue 'P', a letter test match could be a red 'm'); a recombined stimulus (in the example, a letter test match could be a blue 'm'), or a correct feature bound with an extralist feature (in the example, a green 'm'). They found no difference in response times between these three conditions, suggesting that features were not obligatorily bound into objects – if they were, people would be faster when the whole stimulus (the object) was reinstated. Instead, subjects seemed to only retrieve the particular type of feature necessary for accurate performance. Finally, Vergauwe and Cowan (2015) found that the representation of colored shapes could be biased towards a whole-object or a separate-feature representation simply by explicitly instructing the subjects to do so.

There are two possible explanations for these findings: (a) multiple representations are possible – that is, the objects-and-features driven view is correct – or (b) people can flexibly adjust representations during encoding, favoring one type over the other. If the former explanation were true, then these results further suggest that participants can bias their retrieval efforts towards one specific aspect of the representation.

In the present set of three experiments, we aimed to disentangle these two possibilities by investigating the dynamics of retrieval from VSTM. We measured cumulative response-time (RT) distributions for the retrieval of whole objects, and compared these to theorized distributions derived from expectations about the nature of the VSTM representations (more detail on these predictions follows below); these theorized distributions were based on the empirical distributions for single features. (Cumulative RT distributions represent the probability that an RT is less than or equal to some specific time.)

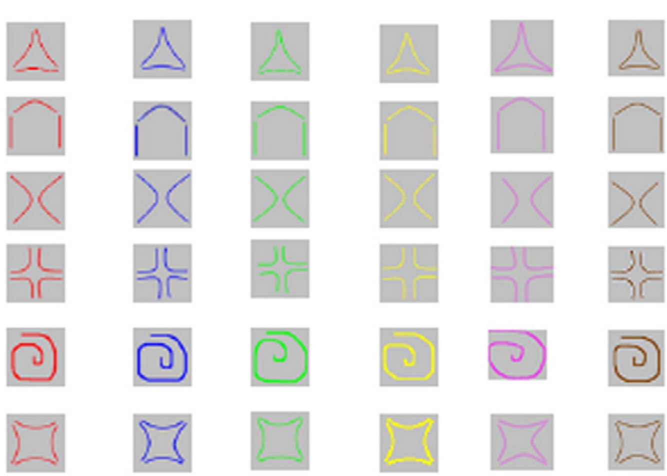

Fig. 1. The 36 colored shapes used in Experiments 1, 2, and 3.

The basic paradigm in our experiments is simple. We showed participants a memory set of three two-featured objects (colored shapes); after a delay, they were probed with a single comparison item – either an individual feature (a shapeless color or a colorless shape, tested in different sessions) or a compound stimulus, that is, the whole object. The crucial manipulation involved the compound-object comparisons. In our first experiment, we forced participants to process the object as a conjunction of features by presenting them with *recombinant lures* (i.e., objects composed of features drawn from different objects in the memory set). In our second experiment, we used *extralist lures*, that is, lures that were constructed from a shape and a color neither of which was presented in the memory set. In this instance, binding of features to objects is completely unnecessary. In our third experiment, both types of lures were combined in an unpredictable manner.

The advantage of recording cumulative RT distributions is that distinct predictions can be made for a feature representation versus a compound-object representation. Experiment 2, with its extralist lures, is the simplest case: In this experiment, we probed memory with a stimulus that is built from two features that were either both present (match probe) or both absent (mismatch probe) in the memory display. If subjects keep features stored independently, a response can be emitted as soon as either of the features can be recognized or rejected. For instance, if the display contained a blue square, a red spiral, and a yellow star, and the match probe is a red spiral, the subject can respond as soon as either the color red or the spiral shape is matched with the memory representation. Conversely, if the mismatch probe is a green disc, the subject can respond as soon as it is clear that neither the color green nor the disc shape match the memory representation. The expected response time can then be modeled as a horse race between the detection times for each of the features – shape or color – separately; within psychology, this model is known as the race model inequality (Miller, 1982; Ulrich, Miller, & Schroeder, 2007). In Experiment 1, the mismatch probe is a recombination of two features from the memory display. (Returning to the previous example, a blue star would be an instance of a recombinant mismatch probe.) If features were kept separate, a response would have to wait at least until both features have been retrieved; it can therefore be no faster than the detection time for the feature that is retrieved most slowly. We are not aware of such a model within psychology, but it reminded us of the process by which the assembly line in a restaurant kitchen operates: A dish can be served only when all its components are ready, and thus the limiting factor is the component that takes the longest time to prepare. Therefore, we will label this model the 'kitchen-line model'.

Both models thus make clear mathematical predictions for the cumulative RT distribution of the compound mismatch probes, based on the cumulative RT distribution of the single-feature mismatch probes. We would expect to observe data in line with the race model inequality in Experiment 2, signifying that features are kept separate; for Experiment 1, we expect that the data will be considerably faster than the kitchen line model prediction if the whole probe was matched to a compound-object representation. The critical comparison is in Experiment 3, where we randomly intermingled recombinant lures (as in Expt. 1) and extralist lures (as in Expt. 2). In this experiment (as in Expt. 1) subjects would need to maintain bound-object representations if they wanted to be able to respond to the recombinant probes correctly. The informative data point would be mismatch probes. If participants would maintain bound-representations only, and extract features from them as needed for the extralist probes, then the results for mismatch probes in Experiment 3 should look like those from Experiment 1 for both types of probes, because the advantage of keeping unbound features available (as in Expt. 2) would have disappeared. If, however, both feature and object representations were to coexist, the recombinant lures would yield results similar to those of Experiment 1, but the extralist lures would conform to those of Experiment 2, because now subjects would have access to unbound features as well.