A compressibility account of the color-sharing bonus in working memory

Hanane Ramzaoui[1] · Fabien Mathy[1]

Abstract
It has been established that objects sharing color in a visual display can boost working memory. The capacity to encode singletons particularly benefits from the repetition of colors encoded as perceptual groups. We manipulated the algorithmic complexity of visual displays to test whether compressibility of information could account for the color-sharing bonus. This study used a free recall working memory task in which the participants were shown displays of 2 to 8 color items. We examined the influence of set size, complexity, number of same-color clusters and amount of color redundancy. The results showed that the probability of correct recall of the pattern and the proportion of similarity between the pattern and the response decreased with an increase of each manipulated variable, except for color redundancy in terms of probability of correct recall. The model performance of complexity did not differ from that of clusters, but complexity was found more accurate than either set size or color redundancy. The results also showed that similar items were more often recalled adjacently, and complexity correlated strongly with the number of extra color repetitions in the response, suggesting that more complex patterns encouraged the use of information compression. Moreover, color repetitions were more often recalled first and the probability of correct recall for singletons and sub-patterns could be predicted by the compressibility measure. We discuss the potential advantage of using compressibility measures to capture the effects of regularities in visual patterns, in particular to refine analysis of the color-sharing bonus.

Keywords Working memory · Visual short-term memory · Perceptual organization · Complexity · Compression

Estimates of working memory capacity vary substantially depending on the type of tasks (Cowan, 2001), type of material (Alvarez & Cavanagh, 2004; Jones & Macken, 2015), number of features (Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013), and training and familiarity with the material (Reder et al., 2016; Simmering et al., 2015). The comprehension of these factors is of paramount importance for determining how objects are encoded in working memory, either as single objects represented in separate slots or depending on a common resource distributed across objects (Awh et al., 2007; Bays et al., 2009; Donkin et al., 2014). For instance, Quinlan and Cohen (2012) concluded that grouping by similarity benefits do not naturally derive from either the simple slot-based account or the resource-limited account.

To better understand how objects are encoded in working memory, there has been an increased interest in investigating how relational information can be structured in working memory (e.g., see Brady et al., 2011b; Gao et al., 2016, Jiang et al., 2000; Jiang et al., 2004) instead of simply focusing on how many basic items can be remembered independently (Miller, 1956). Effectively, another source of variation when estimating capacity comes from the ability to chunk information (Cowan, 2001; Orbán et al., 2008; Norris & Kalm, 2019), for instance when correlated features are introduced over the long term (Brady et al., 2009), when groups form spontaneously (Brady & Tenenbaum, 2013; Mathy & Feldman, 2012), and from the ability to extract a statistical summary of an entire visual display (Brady & Alvarez, 2011a).

One example of how redundant information affects the span is a study by Morey et al. (2015), who showed a boost of working memory capacity when objects shared colors in visual displays. The authors found that the capacity to encode non-duplicate colors benefited from duplicate colors. An idea is that information compression processes

✉ Hanane Ramzaoui
hanane.ramzaoui@univ-cotedazur.fr

[1] Université Côte d'Azur CNRS, BCL, Nice, France

(Brady et al., 2009; Norris et al., 2020; Mathy & Feldman, 2012) could potentially account for the results of Morey et al. (2015) For instance, Chekaf et al. (2018) showed that complexity of the memoranda is a clear indicator of how working memory can succeed in optimizing storage on the spot. We also know that a signature of the compressibility account benefits on singletons (Brady et al., 2009; Norris et al., 2020). The to-be-tested idea initiated by these authors is that more compressed representations in working memory can leave room for remembering extra material. Therefore, it seems crucial to measure concurrently how many duplicates and how many singletons can be recalled by experimentally designing the duplicates (see Benchmark 1.3 in Oberauer et al., 2018).

Morey et al. (2015) tested two accounts of the color-sharing bonus: i) the boost is determined by a reduction of information load due to the shared colors that can be grouped together or ii) automatic attentional capture is induced by the duplicates and this facilitates encoding of the visual display. The authors found that duplicate objects first effectively captured attention, and that this initial step was followed by a greater focus on unique colors. Our general hypothesis is that the observation of Morey et al. (2015) seems in line with a compression account positing that redundant information (i.e., clusters) could be first compressed to leave room for less compressible information (i.e., singletons). Following a compression account, redundant information in display sets containing repeated colors could be compressed in working memory (Chekaf et al., 2016; Mathy et al., 2016). By analogy to what Morey et al. (2015) found, the to-be-compressed objects could first capture attention, and this initial step could be followed by a greater focus on unique colors. If a processing component in working memory can effectively compress information on the spot, it means that optimization of information might prevail over attentional and grouping processes combined to account for the color-sharing bonus phenomenon. The rationale is that a compression account could explain the color-sharing bonus as the result of individuals' attempts to reduce information load in stimulus sets, and this would account for why attentional processes in working memory are first directed towards redundant information (i.e., to best optimize redundant information first). As a result of the compression process, the compressed information could leave room in memory for encoding less compressible information such as singletons. For instance, for the stimulus display blue-blue-blue-red, the newly encoded chunk 'blue-blue-blue' should make the stimulus display easy to recall, with no particular benefit on 'red'. Correct recall of 'blue-blue-blue' should have no influence on the recall of the last color. However, if 'blue-blue-blue' could be compressed as '3blue', the shorter description should benefit the singleton 'red' because '3blue' supposedly would require less storage space. A last advantage of testing the compression account is that compressibility measures are versatile and can capture diverse effects of redundant information on memory (e.g., symmetries, alternating patterns, and clusters).

The present study The purpose of this study was to test whether compressibility of information can offer a potential account of the color-sharing bonus in visual working memory. With this goal in mind, we decided to switch paradigms. Previous studies have largely employed the change detection paradigm (Luck & Vogel, 1997; Rouder et al., 2011). This paradigm uses either single-probe display tests or whole display tests to assess whether participants can detect a change in a set of objects (see Rouder et al., 2008, 2011), but these tests do not properly enable to examine how sets or subsets of items are recalled. In this study, we used a free recall procedure to assess not only the probability of correct recall, but also how groups of items are being formed.

The participants were shown displays from 2 to 8 set sizes with a brief delay, after which they attempted to recall the color of each stimulus item in the order of their choice. We varied the compressibility of the displays to test whether greater compressibility could lead to better memory performance. To manipulate the compressibility of our material, we introduced statistical regularities by clustering parts of the displays using color duplicates. Our protocol allowed us to examine the influence of set size, color redundancy, and number of same-color clusters, but concurrently, we used existing algorithmic complexity measures to test whether compressing information could account for the color-sharing bonus in working memory. The complexity metric is elaborated later in the Method section after we introduce our material.

Method

Participants

Fifty volunteers (25 females, 25 males) aged between 18 and 29 years old ($M = 21.56$, $SD = 2.47$) participated in the study. The participants were psychology students at the Université Côte d'Azur, France. All reported normal or corrected-to-normal visual acuity. None reported having any color vision problems. Informed written consent was obtained from all of the participants.

Apparatus and stimuli

The experiment was coded in HTML and run on a 7.9″ tablet (iPad Mini 2, Apple Inc.) (17.7 × 10 cm). Responses were given manually by touching the sensitive screen.

The stimuli consisted of colored squares. There were four possible colors: blue, red, green, and yellow (see Table 1). These colors were selected, as they are easily distinguished from one another and discriminable with a medium-gray background (RGB value: 50, 50, 50). The size of each square was 1.00 × 1.00 cm, subtending approximately 2° × 2° degrees of visual angle. The squares were placed within a virtual grid made of 12 equally sized columns and 12 rows. The virtual line of the first column started at 0.25 cm of the left edge of the screen and the space between two squares was 0.5 cm. The position of the colored squares on the y-axis could vary in steps of 0.5 cm from 0.25 cm to 8.75 cm of the vertical screen.

For simplicity purposes, we built our visual patterns based on vectors of characters for which we could easily obtain compressibility measures (see below), in contrast to using two-dimensional matrices for which algorithmic complexity cannot be obtained directly (Kempe et al., 2015), in particular when using four symbols (for which there is no metric yet in 2D, see https://algorithmicnature.org). The use of vectors was sufficient to build clusters the following way. For each item, the program attributed a random column from left to right on the x-axis. As an example, for the made-up pattern '333111' (in which the colored squares are here represented by different numbers), each of the six items were randomly attributed a single column, but with the choice of the columns obeying a left-to-right order of the items (for all of the trials). For instance, the two first items '3' were placed on the first two columns, then, one '3' one column away was on the right, and the three last '1' items appeared every two columns. A pattern with the dashes representing empty columns was then '33-3-1-1-1- -'. In sum, our design enabled us to focus on whether duplicate colors were contiguous or not given a one-dimensional horizontal direction (i.e., to probe the formation of groups), with vertical exact locations being randomized across trials and participants.

Complexity measurement

To quantitatively and accurately measure the process of compressibility, we estimated the complexity of our patterns using an approximation of algorithmic/Kolmogorov complexity (Li & Vitányi, 2008). Algorithmic complexity is a notion that was initially developed by Kolmogorov (1965). The algorithmic complexity of a sequence is the length of the shortest program allowing to reconstruct the sequence.

The algorithmic complexity of a string s is more formally defined as the length of the shortest program running on a given universal Turing machine that will produce s and then stop. The algorithmic complexity is known to be incomputable (Li & Vitányi 2008) but it has been more recently suggested that an approximation of algorithmic probability (for short strings in particular) can be obtained[1] by running a large random sample of small deterministic Turing machines (Soler-Toscano et al., 2013, 2014). To obtain the approximation of the Kolmogorov complexity of our patterns, we used the Algorithmic Complexity for Short Strings (ACSS) package (Gauvrit et al., 2015).

The higher the algorithmic complexity (complexity K), the lower the compressibility of a pattern. In more concrete terms, this means that if an ideal GIF lossless compression algorithm existed (this is not the case and this could not be demonstrated since algorithmic complexity is not computable), K (the size of the GIF file) would be the estimate of the best-compressed version of a visual pattern. Here, K is simply the best approximation of the best-compressed version of our vectors.

The complexity of our 28 patterns varied from 5.41 to 27.41 (22 levels of complexity in total). See Table 1 for a sample of patterns and their algorithmic complexity. The complete list of patterns and their corresponding complexity is provided in Supplementary Material.

Procedure

The 28 patterns were predefined and remained the same for all of the participants in order to obtain a satisfying range of complexity values. The choice of colors was randomized for a given trial. For example, if the pattern was '23111', the program randomly assigned a first color such as red to '2', a second color such as blue to '3', and a third color to '1' different from the two first selected colors, such as green. The program then generated the following display of colored squares (from left to right on the screen): red, blue, green, green, green.

The number of stimulus items varied between 2 and 8 colored squares (set size 2, set size 3, set size 4, set size 5, set size 6, set size 7 and set size 8). We also measured how items were clustered within patterns based on feature similarities. Each stimulus display contained one (for set sizes 3 and 4) or more (for set sizes 5 to 8) groups of items of the same color. The amount of color redundancy

---

[1]Technically, this method is equivalent to running a sample of random programs on a universal Turing machine, without relying on a specific choice of machine. It requires a sample of billions of Turing Machines runs on a super computer to determine a reliable approximation of algorithmic complexity.

Table 1 Algorithmic complexity, number of same-color clusters, and amount of color redundancy for a sample of patterns

Set Size	Pattern	Complexity K	Number of clusters	Amount of redundancy
2		5.41	2	0
2		5.41	2	0
3		8.41	2	0.67
3		8.52	3	0.67
4		11.76	3	0.75
4		12.23	4	0.50
5		15.15	4	0.50
5		16.23	5	0.40
6		18.45	4	0.67
6		19.42	5	0.42
7		21.63	6	0.43
7		23.75	7	0.28
8		24.85	5	0.44
8		26.31	6	0.34

within each pattern varied from 2 to 4 items with the same color. Only the smallest set size (set size 2) contained only singletons. Set sizes less than 8 could include unique colors within the display array. As there could be several groups of items sharing the same color in a given pattern, with varying amount of redundancy, we then considered the sum of all redundancies divided per the number of grouped items sharing the same color. To control the influence of set size, the amount of redundancy was also scaled to set size. For example, for the pattern '24412' : the sum of redundancies is 4 and these redundancies were divided into two groups of colors (2 and 4). The amount of redundancy was then $(4 \div 2) \div 5 = 0.4$.

We considered that a set of items could be grouped whenever objects with the same color could be placed next to each other (for example, '23111'), but not when objects were separated by intervening items of other colors (for example, '13241'). For example, the pattern '2121224' contained one redundancy of contiguous items of the same color ('22'), five singletons ('2', '1', '2', '1' and '4') and six clusters: '2 / 1 / 2 / 1 / 22 / 4'. The number of same-color clusters and the amount of color redundancy are reported in Table 1 for a sample of patterns. Note that we consider clusters as a more limited description than algorithmic complexity that is supposed to capture a larger number of regularities such as alternating patterns, symmetries, and clusters. We also measured the number of repetitions in both the stimulus displays and the response. A repetition was considered two adjacent similar items. This will be detailed later in the Data analysis section.

The experiment was conducted individually in a dimly illuminated room. Each participant was required to recall as many colors as possible, by touching the tablet. The tablet was placed on a table at a distance of approximately 30 cm from each participant. The experiment began after the experimenter ensured that each participant had understood the instructions. During the trial, a central fixation cross appeared for 1000 ms on a medium-gray background before the study array. The study array was presented briefly for 200 ms on a medium-gray background, followed by a 1000-ms retention screen. The test screen then displayed the locations of the colored squares replaced by black question marks (the question marks matched both the location and the size of the initial colored squares). The participants were asked to click on the question marks in the order of their choice to recall each of the initial colors. When clicked, one question mark was replaced by a quadrant of four color possibilities on a dark-gray background. The arrangement of the colors in the quadrant randomly changed during each trial to avoid any spatialization strategy during encoding. The participants were instructed to provide the best guess whenever they felt unsure of their responses. The next trial was initiated after the participant clicked on the 'Next' button. Rapid feedback was provided after each trial to keep the participants motivated. A positive feedback 'Bravo!' was displayed whenever the participant recalled correctly all of the items. The feedback 'Not exactly...' was displayed when at least one error was committed. See Fig. 1 for the timeline of a trial. The 28 patterns were presented in random order and only once to each participant. The number of trials
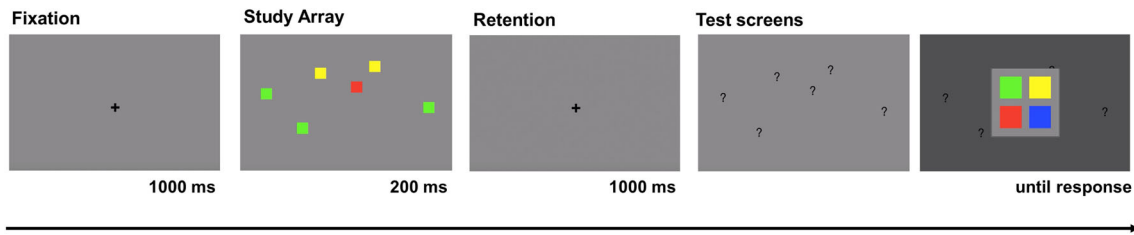
Fig. 1 Example of one trial and procedure. *Note.* The sequence of screenshots depicts one trial for set size 6, for which the algorithmic complexity was $k = 19.42$. After presenting the stimuli, the participants were instructed to click on all of the question marks in turn, in the order they preferred, to recall the colors. For each item, a choice was voluntarily not too high to study a spontaneous recoding process, that is, under conditions that limited any training effect (the experiment lasted approximately 10 min).

of colors was displayed in a random quadrant after the question mark at the item's location was clicked on. A new trial was initiated after a feedback was provided to the participant. See the online version of the paper for a colored version of this figure

Data analysis

We measured for each trial whether the pattern of response was correct. Only perfect recall of a pattern was scored correct and coded 1 (vs. 0) when the participant perfectly recalled all of the items (i.e., all colors of a pattern correctly recalled within a trial). Trials with unanswered items were excluded from analysis (approximately 1.5% of the trials).

A second dependent variable was the proportion of similarity between the pattern and the response for each trial. The similarity was calculated using the optimal string alignment method from the R package StringDist (van der Loo, 2016). The distance was first calculated by counting the number of deletions, insertions, and substitutions necessary to turn the response into the original pattern of the stimulus display. The distance was then divided by the maximal possible distance for a given length. This proportion was finally subtracted from 1 to obtain a similarity measure between 0 and 1, with 1 corresponding to perfect similarity (distance 0) and 0 to complete dissimilarity. A last dependent variable was the response time.

Measures of WM and complexity To analyze the respective effect of set size, amount of color redundancy, number of clusters, and complexity K, statistical analyses were conducted using generalized linear modeling (GLM) as implemented by the R package lme4 (Bates et al., 2015). Analyses were performed using R version 3.6.2 (The R Foundation's Project for Statistical Computing). Four sets of separate models were evaluated. A binomial logistic regression was conducted to investigate the influence of Set size in predicting the probability of correct recall of the entire pattern. A separate model was used to investigate the influence of the amount of color redundancy in predicting the probability of correct recall of a pattern. Another separate model was used to investigate the influence of the number of clusters in predicting the probability of correct recall of a pattern. A final model was used to investigate the influence of complexity K in predicting the probability of correct recall of a pattern. Because complexity K reflects set size, additional analyses were carried out to investigate the influence of complexity K for constant set sizes (expect for set size 2, as the level of complexity K was the same for all patterns in this condition). To predict the proportion of similarity between the pattern and the response from our set of continuous predictor variables, we used quasibinomial models with the same model structures as for the probability of correct recall of the patterns. Only p values below the .05 threshold were considered noteworthy. Adjusted R-squared values were also reported to indicate size effects.

Note that the correlations between complexity K, set size and number of clusters all significantly correlated positively ($r \geq .85$, $p < .001$; see Table 2). To refine the analysis of the respective effect of number of clusters and complexity K, statistical analyses on response times

Table 2 Correlations between complexity K, set size and number of clusters

Variable	n	M	SD	1	2	3
1. Complexity K	22	15.43	6.88	—		
2. Set size	8	4.96	1.99	.99***	—	
3. Number of clusters	7	4.08	1.60	.88***	.85***	—

***$p < .001$

were conducted using a linear model (LM). Two sets of separate models were evaluated. A linear regression was conducted to investigate the influence of number of clusters on response times for recalling a pattern. A separate model was used to investigate the influence of complexity K on response times for recalling a pattern. For this analysis, we only considered trials in which the participants perfectly recalled all of the items. Then, we excluded the trials for which response times were longer than 3 SD above the mean (these outliers were approximately 1.6% of the trials). Response times presented skewed distributions and were thus log-transformed to meet the LM assumptions. After this selection, it remained 20 levels of complexity K (initially 22 levels), ranging from 5.41 to 26.31. As for number of clusters, it remained 5 levels (initially up to 8 clusters), ranging from 2 to 6.

Model performance ROC curves were used to build six models based on either (1) set size, (2) amount of redundancy, (3) number of clusters, (4) complexity K, (5) set size + amount of redundancy and (6) set size + number of clusters. The area under the roc curve (AUC) value is a score between 0 and 1. This area represents the model performance. AUC quantifies how much a model is capable of separating correct responses from incorrect responses. The ROC plots are shown in Fig. 4. AUC close to 1 indicates that the model fully explained the data and a value of .50 indicates a model performance not better than chance level. The AUC values and confidence intervals were calculated using the R package pROC (Robin et al., 2011). All of the AUC values were computed with 95% intervals and were compared using the Delong method (DeLong et al., 1988) with the roc.test function.

Memory compression This subsection more specifically focused on the analyses of sub-patterns, color repetitions and on how items were ordered in each of the participants' responses. To refine the effect of information compression on performance, a quasibinomial logistic regression was first conducted to investigate the influence of complexity K on the correct recall for singletons. This analysis aimed to test whether simpler patterns could leave room in memory (for instance, for encoding the singletons of a display) than more complex patterns. We then carried out another quasibinomial logistic regression with the number of clusters as a predictor of the correct recall for singletons. To further assess the compression effects, we analyzed memory performance for eight-item patterns, with a test of the influence of complexity K for the first items predicting the probability of correct recall for the rest of the items. This analysis targeted whether more compressed representations of one part of a pattern could leave more room for encoding

the second part of the pattern (thus, it made sense only for the longest stimuli).

Beforehand, we conducted an analysis to test whether there was a trend toward a left-to-right recall order in the participants' responses. To do this, we analyzed the order in which participants clicked on the items during the recall phase. For each pattern, we took all pairs of items from the participant's response and we summed the differences between their x positions (x refers to the column number in which the item was displayed). A left-right order of response resulted in a negative value. We then conducted a one-sample t test (one-tailed) to analyze whether the mean distance between the recalled items was significantly lesser than 0, for each set size. The mean distance between recalled items was significantly lesser than 0 for set sizes 1 to 4, all $t \leq -3.08$, all $p < .001$, all $d \geq 0.21$, but it did not differ from 0 for set sizes 5 to 8, all $t < 1$, all $p \leq .347$. These results suggest a tendency to scan and recall items from left to right when the set size was limited. Beyond set size 4, more complex grouping processes might have hindered this tendency. To refine this analysis, we tested whether there was a trend toward a left-right encoding strategy against the opposite right encoding strategy, based on the similarity between the participant's response and the stimulus pattern. For this purpose, we first computed the similarity between the response in the order of item recall (e.g., 4113) and either the pattern in its left-right order (e.g., 1413), either the pattern in its (reverse) right-left order (3141). We then conducted a two-independent samples t test (one-tailed) to compare these two similarities, for each set size (only for trials in which the participant perfectly recalled all items). The mean similarity between the response and the pattern in its left-right order was significantly greater for the reverse order of the pattern for set size 2, set size 3 and set size 7, $t \leq 6.44$, all $p < .001$, all $d \geq 0.31$. However, although there was a trend, the test did not reach significance for set size 4, set size 5, set size 6 and set size 8, all $t \leq 1.35$, all $p \geq .096$.

Note that the tendency to recall from left to right does not necessarily strictly corresponds to the tendency to encode from left to right. Independently, we conducted another analysis to show that participants who were given the opportunity to easily encode the first half of the stimulus (i.e., the left part) performed better for the second half. One example is a participant who could benefit from a cluster of two red items in the stimulus pattern 'red-red-blue-green'. We predicted better recall of 'blue-green' once 'red-red' encoded. However, this did not mean that 'red-red' should be recalled first by the participant. One possibility is that some participants choose to recall some of the most difficult items first, for instance blue and green, followed by the cluster in-mind 'red-red', which might consume less resource.

We then estimated the most probable span based on our compressibility measure. To achieve this, we first computed the local complexities of 8-item patterns based on the local complexity() function of the R package ACSS (Gauvrit et al., 2015), which returns the complexity of sub-patterns. We computed the local complexity with a sliding window of sub-patterns with length span ranging from 3 to 8 items. Four symbols were entered (Alphabet = 4), as the patterns were composed of 3 and 4 colors. For example, the local complexity of the pattern '12232113' with span = 6 returns $K_4(122321)$, $K_4(223211)$ and $K_4(232113)$, which equals 19.40, 18.98, and 19.26, respectively. The pattern '12232113' with span 6 gives a mean local complexity of 19.21. The same pattern gives the mean local complexity of 15.57 with span 5. Note that when no indication is given, for instance in Table 1, it means that K was computed with a maximum window corresponding to the object's length.

In order to estimate the span, we built a linear regression model to predict the similarity between the pattern and the response based on span 3, span 4, span 5, span 6, span 7, and span 8 as predictors. We compared the models using the QAIC function of the R package MuMIn (Barton & Barton, 2015). The criteria for model selection in QAIC (Quasi-AIC) is a modification of Akaike information criterion (AIC, Akaike, 1974) for overdispersed count data. The most parsimonious model is the one that has the lowest QAIC. The model selection for these subspans (see Table 3) indicated that the best-fitting model with the lowest QAIC is the one with span 6 as a predictor of the similarity between the pattern and the response. The QAICs of the span 4 model and of the span 5 model were nevertheless very close to those of the span 6 model (QAIC of span 4 = 225.8, QAIC of span 5 = 225.9 and QAIC of span 6 = 225.9). We can therefore consider that spans 4, 5, and 6 are the best predictors of the proportion of similarity between the patterns and the responses, suggesting that when asked to recall eight-item patterns, the participants tended to encode sub-patterns with either six items, five items, or four items. These results may exceed the suggested capacity limit of working memory of four chunks (Cowan, 2001), but the presence of color redundancies in the patterns can simply explain the capacity to encode more than four items at a glance.

According to the estimated spans (span 6, span 5, and span 4), we built three sets of separate models investigating the influence of complexity K of the first items in predicting the probability of correct recall for the last items. A binomial logistic regression was conducted to investigate the influence of complexity K of the six first items in predicting the probability of correct recall for the two last items (same reasoning for 4-4 and 5-3, instead of 6-2). For example, considering span 6, in the pattern '12232113', we examined if greater compressibility of the first sub-pattern of six items ('122321') could leave room for a better retention of the second sub-pattern of the two last items ('13'). Following the same reasoning, we built three sets of separate binomial regression models to investigate the influence of the number of clusters within the first items in predicting the probability of correct recall for the last items. Considering span 6, we analyzed the influence of the number of clusters within the first six items in predicting the probability of correct recall of the last two items (same reasoning for 5-3 and 4-4).

We further examined the mechanisms underlying the formation of clusters in mind, with the idea that participants could take profit of regularities to simplify the recall process. To test whether participants tended to recall similar items adjacently, we coded the presence of adjacent color items in both the stimulus pattern and in the response (trial-by-trial) and we summed the adjacent repetitions of colors. For instance, for a pattern '1121', we considered that the pattern included only one repetition because '1' was repeated once in '11'. For a response such as '1112', the sum of color repetitions was equal to 2. We then computed the difference between the number of repetitions in the response and the number of repetitions in the stimulus pattern. For a pattern '11221122' (four adjacent repeated items) and a response '11112222' (six adjacent repeated items), the difference would have been 2. This measure was used to estimate how many items were regrouped by

Table 3 Model selection for the sub-span analysis

Model	K	QAIC	Δ QAIC	Akaike weight
Span 6	2	225.8	0.00	0.064
Span 4	2	225.9	0.09	0.061
Span 5	2	225.9	0.10	0.061
Span 8	2	226.0	0.24	0.057
Span 7	2	226.1	0.33	0.054
Span 3	2	230.5	4.69	0.006

Note. K refers to the number of parameters in the model

the participant in comparison to the left-right organization of the pattern. The hypothesis was that the presence of extra repetitions in the response would provide window for clustering. We then conducted an independent samples t test (two-tailed) to analyze whether the sum of repetitions in the response was significantly greater than the sum of repetitions in the stimulus pattern, for each set size (not counting set size 2, which contained only singletons). We selected trials in which the participant recalled perfectly all of the items. In addition, we tested whether more complex patterns encouraged the use of information compression, by performing a correlation between complexity K and the number of extra repetitions in the response.

Finally, we tested the hypothesis that color repetitions should be more frequently recalled first (than last) in the response. For this purpose, we also counted the number of repetitions as a function of their position in the response: first, middle, and last. We considered a repetition position as occurring 'first' when two adjacent similar items appeared first in the response, while we coded 'last' two adjacent similar items appearing last in the response. We coded a repetition position 'middle' otherwise. To control the influence of set size, we computed the proportion of repetitions appearing in the middle position (for each trial), by dividing the total number of repetitions in the middle position in the response per the maximum possible repetitions allowed by the considered pattern. For example, considering the response '133222', $N_{repetitions-in-First-pos} = 0$; $N_{repetitions-in-Last-pos} = 0$, and $Prop_{repetitions-in-Middle-pos} = 1$. We obtained 1 for the middle positions because of the two summed repetitions in the middle, and based on a maximum of two repetitions in our patterns of set size 6, hence $2/2 = 1$ (note that our patterns of set size 6 would not allow 3 repetitions in the middle, which could have been the case if they had been made of a single color). We then conducted a two paired samples t test (one-tailed) to analyze whether the proportion of repetitions in the first position in the response was significantly greater than those in the last position. We also conducted two paired samples t tests (one-tailed) to analyze whether the proportion of repetitions in the first position or last position in the response was significantly greater than those in the middle position. Note that this analysis was restricted to correct response trials only, and we also excluded set size 3 as this condition only allowed repetitions in first and last positions.

Results

Measures of WM and complexity Figure 2 shows memory performance as a function of set size and number of clusters. As expected, the results showed that memory performance

decreased with increased set size $[(\beta = -0.823, z = -19.28, p < .001, R^2_{Adjusted} = 0.37)$ for the probability of correct recall of the entire pattern and $(\beta = -0.438, t = -19.31, p < .001, R^2_{Adjusted} = 0.30)$ for the proportion of similarity between the pattern and the response]. We also found that memory performance decreased with increased number of clusters per pattern $[(\beta = -1.171, z = -19.14, p < .001, R^2_{Adjusted} = 0.40)$ for the probability of correct recall of the entire pattern and $(\beta = -0.558, t = -21.65, p < .001, R^2_{Adjusted} = 0.35)$ for the proportion of similarity between the pattern and the response]. The effect of amount of color redundancy did not reach significance for the probability of correct recall of the entire pattern $(\beta = 0.421, z = 1.67, p = .095)$, but it reached significance when considering the proportion of similarity between the pattern and the response $(\beta = 0.738, t = 4.13, p < .001, R^2_{Adjusted} = 0.01)$. The results also showed increased response time with increased number of clusters, $\beta = 0.108, t = 24.81, p < .001, R^2_{Adjusted} = 0.47$.

As shown in Fig. 3, our results also revealed that memory performance decreased with increased complexity K values $[(\beta = -0.249, z = -19.60, p < .001, R^2_{Adjusted} = 0.39)$ for correct recall of the entire pattern and $(\beta = -0.133, t = -20.16, p < .001, R^2_{Adjusted} = 0.33)$ for the proportion of similarity between the pattern and the response]. The results also showed increased response time with increased complexity K, $\beta = 0.026, t = 32.70, p < .001, R^2_{Adjusted} = 0.61$.

Additional analyses showed that memory performance decreased with increased complexity K for set size 4, 5, 6, 7, or 8 $[(all \beta \leq -1.323, all z \leq -2.30, all p \leq .021, all R^2_{Adjusted} \geq 0.04)$ for correct recall of the entire pattern and $(all \beta \leq -0.332, all t \leq -3.20, all p \leq .002, all R^2_{Adjusted} \geq 0.04)$ for the proportion of similarity between the pattern and the response]. However, the same test did not reached significance for set size 3 $[(\beta = 1.046, z < 1, p = 0.810)$ for the probability of correct recall of the entire pattern and $(\beta = 0.887, t < 1, p = 0.835)$ for proportion of similarity between the pattern and the response]. This absence of a significant effect for set size 3 was likely due to the close levels of complexity. Indeed, each of the set size 3 patterns were characterized by a complexity K of 8.41 or 8.52.

Model performance Figure 4 shows the AUC values for six different models predicting the probability of correct recall of the entire pattern. The six models were a logistic regression based on (1) set size, (2) amount of redundancy, (3) number of clusters, (4) complexity K, and a multiple linear regression based on (5) set size and amount of redundancy and (6) set size and number of clusters. DeLong's test was used for comparing AUC of their ROC curves.
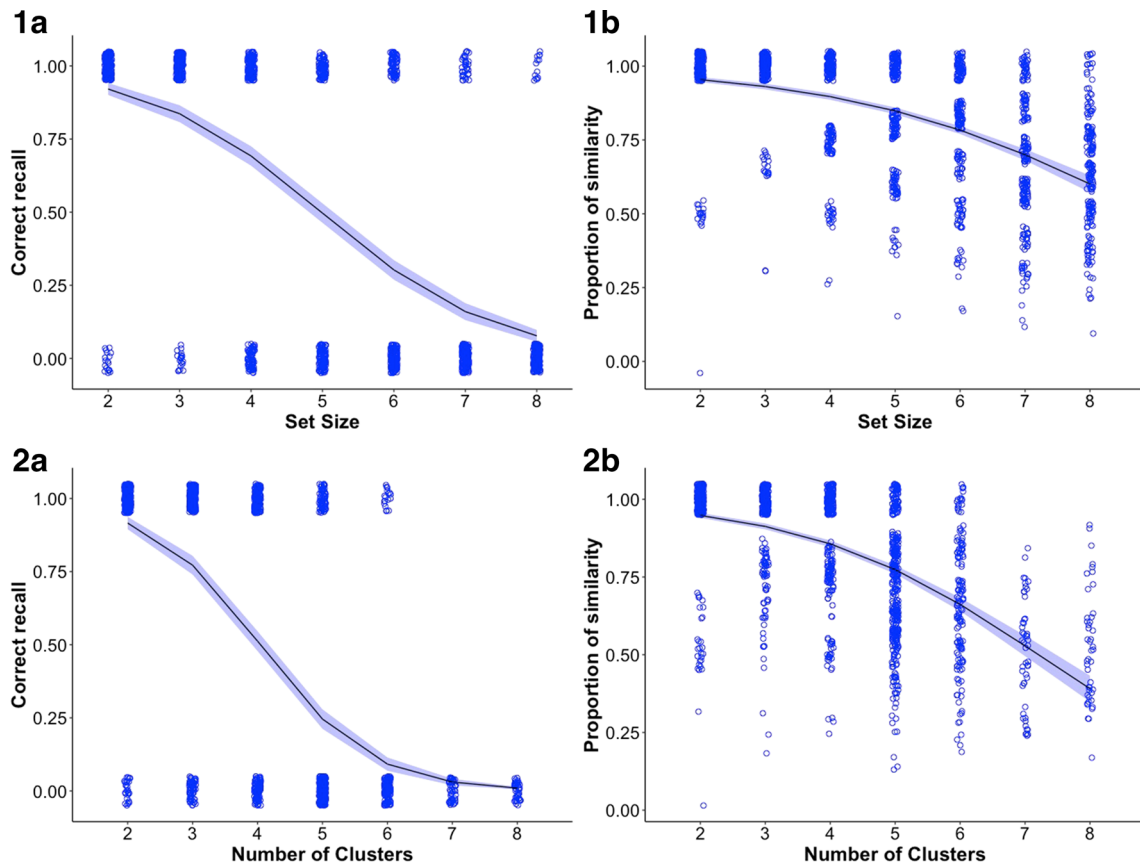
Fig. 2 Correct recall of the entire pattern for a given trial and proportion of similarity between the pattern and the response as a function of set size (1a & 1b) and number of (color) clusters (2a & 2b). The *blue regression line* is fitted to all of the data points and the *light blue area* depicts 95% confidence intervals
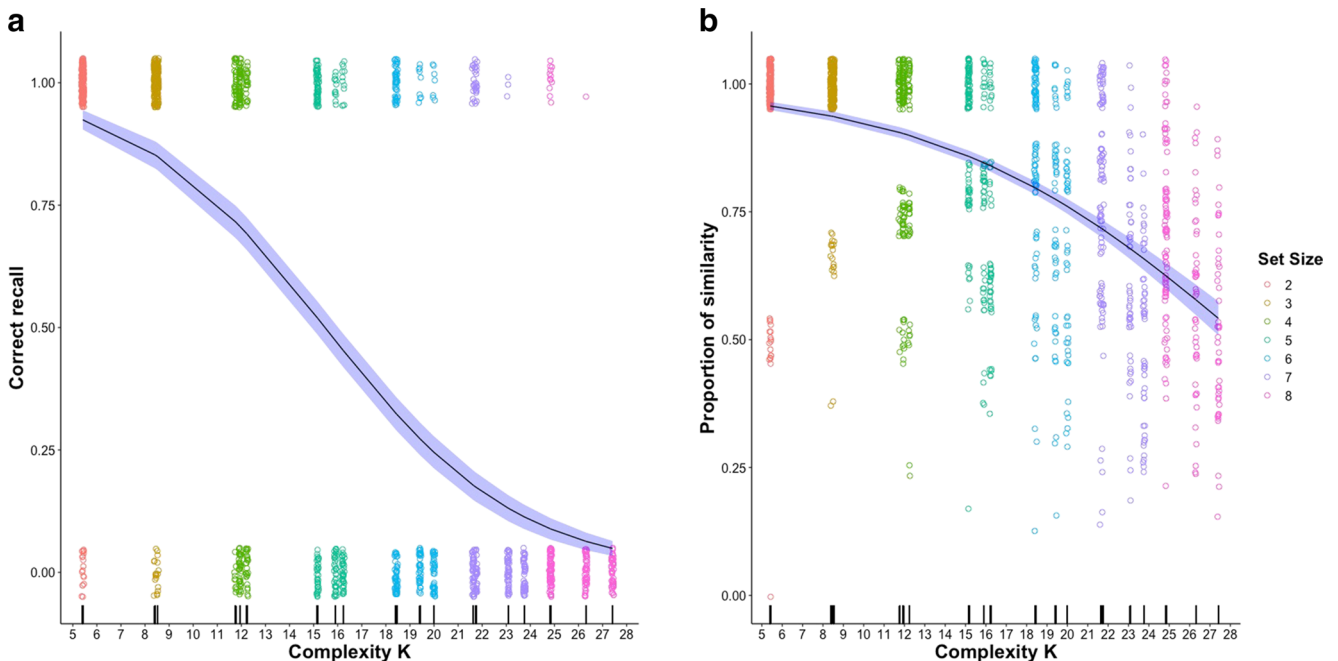
Fig. 3 Correct recall of the entire pattern for a given trial (a) and proportion of similarity between the pattern and the response (b) as a function of complexity *K*. The *blue regression line* is fitted to all of the data points and the *light blue area* depicts 95% confidence intervals. The black ticks on the *x*-axis depict the 22 levels of complexity. See the online version of the paper for a colored version of this figure
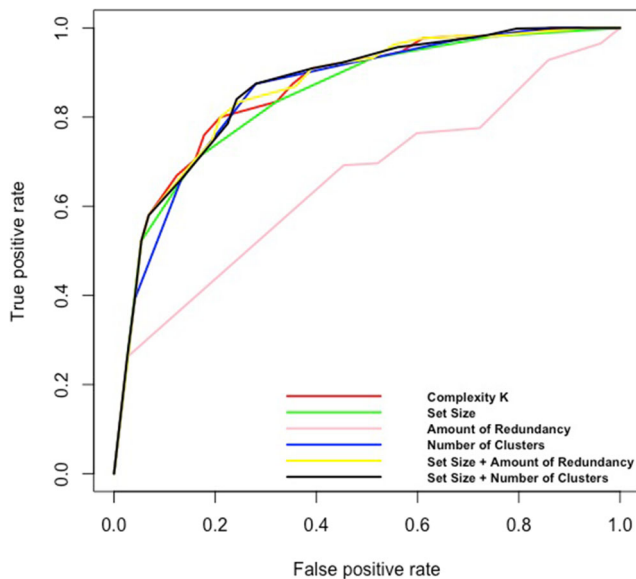
Fig. 4 Receiver operating curves for six models predicting the participants' responses: complexity K model, set size model, number of (color) clusters model, amount of (color) redundancy model, set size + amount of redundancy model, and set size + number of clusters model. *Note*. AUC of complexity K model: 0.862. AUC of set size model: 0.847. AUC of amount of redundancy model: 0.657. AUC of number of clusters model: 0.857. AUC of set size + amount of redundancy model: 0.863. AUC of set size + number of clusters model: 0.863. See the online version of the paper for a colored version of this figure

The result was that complexity K (Model 4), set size and amount of redundancy (Model 5) and set size and number of clusters (Model 6) equally accurately separated correct from incorrect participants' responses, with no significant difference between their AUCs (AUC of 0.862 vs. AUC of 0.863 vs. AUC of 0.863, respectively), all $z < 1$, $p \geq$ .396. The model based on amount of redundancy alone (Model 2) poorly predicted the participants' responses (AUC of 0.657), and its AUC differed significantly from that of complexity K (AUC of 0.657 vs. AUC of 0.862, respectively, $z = 12.73$, $p < .001$) and number of clusters alone (AUC of 0.657 vs. AUC of 0.857, respectively, $z = 12.93$, $p < .001$). There was no significant difference between set size (Model 1) and number of clusters (Model 3) in their AUCs (AUC of 0.847 vs. AUC of 0.857, respectively), $z = 1.43$, $p = 0.152$, but there was a significant difference between set size (Model 1) and amount of redundancy (Model 2) in their AUCs (AUC of 0.847 vs. AUC of 0.657, respectively, $z = 11.59$, $p < .001$). Finally, the model based on complexity K alone did not differ from the model based on number of clusters alone (AUC of 0.862 vs. AUC of 0.857, respectively), $z < 1$, $p = 0.324$, but it was found more accurate than the model based on set size alone (AUC of 0.862 vs. AUC of 0.847, respectively), $z = 9.25$, $p < .001$.

Memory compression We now more specifically focus on the effect of sub-patterns and color repetitions. The previous analysis based on AUC values let appear that several predictors can offer good predictions of memory performance. We then test the more specific hypothesis that simpler patterns should leave more room in memory for singletons. The results showed that correct recall for singletons decreased with increased complexity K, $\beta = -0.140$, $t = -17.66$, $p < .001$, $R^2_{Adjusted} = 0.26$. Figure 5 shows memory performance for singletons as a function of complexity K. A similar test based on the number of clusters offered a similar pattern of significance, $\beta = -0.554$, $t = -16.77$, $p < .001$, $R^2_{Adjusted} = 0.24$.

Regarding the hypothesis that greater compressibility of the sub-patterns should leave room in memory for the rest of the patterns within trials, our analyses of the eight-item long patterns showed that the probability of correct recall for the last items increased significantly with decreased complexity K of the first items, and this considering all estimated spans (span 6, span 5, and span 4). Indeed, the probability of correct recall for the last two items increased significantly with decreased complexity K of the first six items, $\beta = -2.735$, $z = -6.06$, $p < .001$, $R^2_{Adjusted} = 0.23$. Similarly, the probability of correct recall for the last three items increased also significantly with decreased complexity K of the first five items, $\beta = -4.682$, $z = -4.21$, $p < .001$, $R^2_{Adjusted} = 0.14$. Finally, the probability of correct recall for the last four items also increased significantly with decreased complexity K of the first four items, $\beta = -3.718$, $z = -4.00$, $p < .001$, $R^2_{Adjusted} = 0.08$. Figure 5 shows memory performance for the last two items as a function of complexity K of the first six items.

Similar analyses based on the number of clusters instead of complexity K showed that the probability of correct recall for the last two items increased significantly with decreased number of clusters within the six first items, $\beta = -1.298$, $z = -5.09$, $p < .001$, $R^2_{Adjusted} = 0.15$, and also that the probability of correct recall for the last three items increased significantly with decreased number of clusters within the five first items, $\beta = -0.729$, $z = -2.99$, $p < .01$, $R^2_{Adjusted} = 0.04$. However, the analysis considering the number of clusters within the first four items as a predictor of probability of correct recall for last four items resulted in a non-significant result, $\beta = -0.013$, $z < 1$, $p = .968$.

Also, we found that the sum of repetitions in the response was significantly greater than those in the stimulus pattern for each set size, all $t \leq 6.23$, all $p < .001$, all $d \geq 1.23$. Moreover, complexity K and the number of extra similar adjacent items in the response (i.e., the difference between the repetitions in the response and those in the stimulus pattern) were found to be strongly positively correlated, $r(495) = .65$, $p < .001$. Concerning the hypothesis that
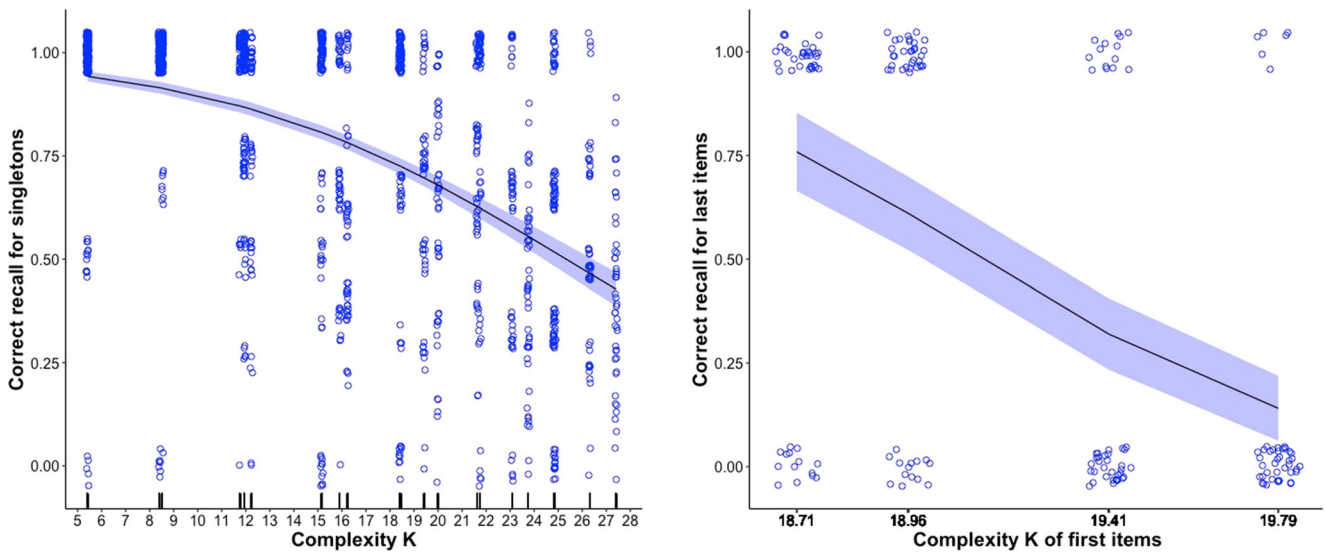
Fig. 5 *Left*: Correct recall for singletons as a function of complexity *K*. The *black ticks* on the *x*-axis depict the 22 levels of complexity. *Right*: Correct recall for the last two items as a function of complexity *K* of the first six items. Note that in both plots, the *blue regression line* was fitted to all of the data points and the *light blue area* depicts 95% confidence intervals
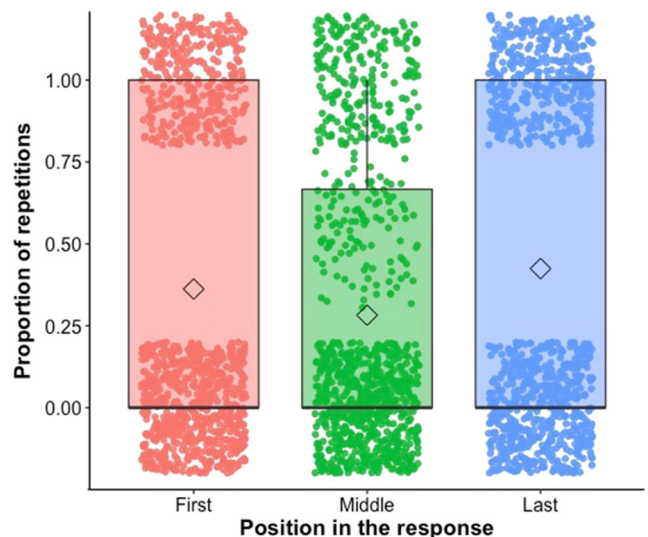
repetitions should be most often observed first in the response, our analyses showed that the mean proportion of repetitions in first position in the response was significantly greater than those in middle position in the response, $t \leq 5.02$, $p < .001$, $d = 0.38$, as well as greater than those in last position in the response, $t \leq 5.21$, $p < .001$, $d = 0.37$. In addition, the mean proportion of repetitions in last position in the response was significantly greater than those in middle position in the response, $t \leq -2.20$, $p < .05$, $d = 0.29$. Figure 6 shows the distribution of the average proportion of repetitions in the responses (by participant and by pattern) for the first, middle, and last positions in trials with repetitions.

Discussion

It is well established that the presence of objects sharing the same color in a visual display boosts memory capacity (Morey et al., 2015; Morey, 2019; Peterson & Berryhill, 2013; Quinlan & Cohen, 2012). The purpose of this study was to use existing objective (algorithmic/Kolmogorov) complexity measures to test a compression account of the color-sharing bonus in visual working memory (Pothos & Chater, 2002). The novelty was that this method can be automated to avoid any assumptions regarding clustering techniques or similarity-based metrics.

The compression account involves that the representation of memoranda is modified to gain storage space. A lossless compression process is theoretically expected to eventually redescribe an object in a shorter way. On the contrary, grouping or chunking processes do not imply that the

pieces of information they integrate have a different nature than in the original object. The main hypothesis was thus that compressed information should leave room in working memory. But beyond that, the presumed conceptual link between compression and memory optimization was that the presence of different forms of regularity (clusters, alternating patterns, symmetries) in the material can help minimize the code length of information in memory. To manipulate the compressibility of patterns of items, we

Fig. 6 Distribution of the average proportion of repetitions in the responses (by participant and by pattern) for the first, middle, and last positions of the items in the response. The *black diamonds* indicate the means. The data points are jittered. Note that a repetition was considered two adjacent similar items. See the online version of the paper for a colored version of this figure

introduced statistical regularities in the stimulus displays by clustering the items horizontally based on feature (color) similarity. We therefore expected the algorithmic complexity factor K to be more predictive than the sole number of clusters in the material. Our results overall only showed a small advantage of K, which seemed to be more sensitive than the number of clusters, for instance to predict that the probability of correct recall for the last four items increase with decreased complexity K of the first four items. The factors are not incompatible, and our results tend to show that if compression occurs, it could mainly be linked to the detection of clusters.

More specifically, we found that memory performance decreased with an increase in number of clusters. This finding suggests that items are not treated (encoded) independently and that observers can easily extract sub-patterns in a visual display. This is in line with behavioral evidence showing that perceptual organization influences the storage of information in visual working memory (e.g., Peterson and Berryhill (2013), Quinlan and Cohen (2012), Woodman et al. (2003), and Xu (2006), Xu and Chun, 2007). This finding also supports a recent study suggesting that clusters of similar colors help increase precision in visual working memory (Son et al., 2020).

Memory compression

We tested whether the capacity to encode uniquely colored objects (singletons) could benefit from the repetition of colors in a visual display, as it has been suggested that attention is particularly engaged in recovering the singletons in the presence of duplicates by reducing the memory load (Morey et al., 2015; Morey, 2019). For instance, Morey et al. (2015) found that attention was initially captured by redundant objects, and then was more focused on single-ton objects during retention. Our hypothesis was that these previous observations could be in line with a compression account positing that redundant information (e.g., clusters) could be first compressed to leave room for less compressible information (i.e., singletons). Consistently, we found that similar items were more often recalled adjacently in the response than in their original form in the pattern and that these color repetitions were more often recalled first than last. Interestingly, complexity K and the number of extra similar adjacent items in the response correlated strongly, suggesting that more complex patterns encouraged the use of information compression (i.e., the participants tended to group more items when the stimulus pattern was more complex). Our findings also show that memory performance for both the singletons within the visual displays and the whole visual displays increased with the compressibility of the patterns (estimated by their underlying algorithmic complexity). Moreover, the compressibility effect was present

across set sizes (from 4 to 8). These results suggest that lower compressibility led to greater memory performance overall, in accordance with previous studies on working memory capacity based on other types of material such as digits (Mathy & Feldman, 2012), multi-dimensional shapes (Chekaf et al., 2016), or previous studies on categorization showing higher categorization performance when information could be minimized (Feldman, 2000; Lafond et al., 2007).

We show that compression can potentially account for the color-sharing bonus in visual working memory, with a strong link to clustering effects. Still, it is not clear how compression processes operate cognitively or perceptually in our task, even though we know that complexity minimization plays a central role in human perceptual organization (Feldman, 2003, 1999; Pothos & Chater, 2002). A few recent models that have proven useful to fit human data in the visual working memory domain (Brady & Tenenbaum, 2013; Orbán et al., 2008; Orhan & Jacobs, 2013; Son et al., 2020) could offer an account of the mechanisms at play, but those formal models do not directly consider the compression hypothesis. In the present study, algorithmic complexity alone revealed to be a more direct predictor of performance than set size and number of clusters together (which still offered good fit of our data). This algorithmic complexity metric conveniently captures many kinds of regularities. Aside exploiting repetitions and symmetries, the algorithmic complexity metric presents global properties, such as being sensitive to set size and number of different features (here, colors) within a set. The logic is that both smaller set sizes and smaller number of different colors increase chance to achieve shorter description of the original object. The capacity of algorithmic complexity to extract many types of structures present in a visual display to express them as the shortest expression may account for our findings.

The algorithmic complexity approach can still appear elusive as remains uncertain how information can be compressed at a psychological or neuronal level, but the metric has been described as universal (e.g., general enough to reflect different languages attempting to compress information). Following Feldman (2003), a fundamental idea is the principle of parsimony stating that simpler ideas are more truthful. Effectively, the simplest representation is more likely to be correct because it offers the right amount of complexity. It is a very old idea developed by William of Occam and later developed mathematically by Rissanen (1978) and Jeffreys (1939) who respectively developed the minimum description length approach and Bayesian inference by penalizing complexity. A basic idea is that a too simple description of an object might lack capturing its essence while a too complex description might rather capture noise in the data (Pitt & Myung, 2002). The

simplicity principle in psychology thus posit that individuals are better off drawing the simplest interpretations to interpret their environment.

Our results also showed that both increased number of clusters and increased complexity K slowed down the recall process (i.e., we observed higher response times on average for reconstructing the entire display set). These findings could suggest that a decompression process could be at play (positing a compression process). Effectively, we could interpret faster responses as reflecting faster decompression of a shorter compressed representation, but a more classic explanation could be that lower memory load produced by simpler displays produces faster response times.

Regarding memory benefits of simpler patterns, analyses of compression effects concerning eight-item long patterns showed better memory performance for the last items when the first items were more compressible. This result suggests that more compressed representations of one part of a visual pattern could have left more room for the encoding of subsequent items (confirming Brady et al., 2009; Reder et al., 2016). This finding is in line with a strategic allocation of attention in memory displays containing repetition of colors, and provides further support for our hypothesis that compressibility may account for the color-sharing bonus in working memory.

However, our analysis may be limited in that our simple design was based on one-dimensional clusters because there is no complexity calculator available yet for 2-D matrices of non-binary symbols (see http://www.complexitycalculator.com). This design let us assume that observers could be influenced by a left-to-right encoding of the items. This assumption was nevertheless motivated by previous studies suggesting that individuals show a spatial leftward bias for visual features in short-term memory (Della Sala et al., 2010), and a leftward bias direction of the first eye movement when encoding visual scenes (e.g., Nuthmann & Matthias, 2014; Dickinson & Intraub, 2009). We did not use an eyetracker to assess this bias as we cannot posit the hypothesis that such a bias would be systematic enough across trials within and between participants. We therefore admit that our metric is only an approximation of the regularity of the spatial configurations of the displays, and as such it cannot be thought to reflect sequential encoding processes. This might explain why the effect size of the analyses concerning the "eight-item long patterns was weak. We believe our metric could help detect trends in the data, but cannot offer a priori perfect fit of all possible encoding strategies.

Model performance

In the visual working memory domain, previous studies have indicated that performance was higher for grouped items than for ungrouped items (e.g., Woodman et al., 2003; Xu, 2006, 2002). In particular, Peterson and Berryhill (2013) reached this conclusion by manipulating the spatial proximity between items sharing the same color, as we did in our study. Our results showing that cluster-based measures perform better than redundancy values are consistent with this literature, suggesting that perceptual grouping per similarity alone is not as powerful as similarity and proximity combined. Interestingly, we found no significant difference between the complexity/compressibility model and the number of clusters model, both models being similarly predictive of memory performance overall. This might be due in part to a non-systematic left-to-right serial encoding bias mentioned above. However, we believe that the good fit of the complexity model is due to its capacity to consider both redundancy and regularity in the visual patterns, which might globally account for the color-sharing bonus in memory in our study. Moreover, even if the stimulus items were not perfectly aligned in the visual display (as we allowed some randomness in the vertical position of each item of a pattern), the complexity/compressibility model was powerful enough to predict memory performance and could make a more powerful prediction than the number of clusters to predict performance for the last four items based on the description of the first four items in eight-item patterns. Our findings therefore indicate that algorithmic complexity alone can be a relevant account of visual working memory online processes. They also suggest that algorithmic complexity can serve as a quantitative and objective measure of the perceptual organization of patterns based on both feature similarity and proximity. Proximity, compared to similarity, has been found to facilitate discrimination (e.g., Quinlan & Wilton 1998; Han et al., 1999, Ben-Av & Sagi,1995), and the combination of similarity and proximity has also been found to boost even more memory performance (Kubovy & Van Den Berg, 2008), but algorithmic complexity alone could help represent different forms of regularities from which observers benefit strongly in discrimination tasks.

In our study, the effect of amount of color redundancy was not found to be the best predictive factor. One explanation can be based on the specificity of our experimental design. For example, Morey (2019) used displays of set size four to six items with either only unique items, either two color duplicates or three color duplicates. In our study, as we varied the set size of our patterns from 2 to 8, we had only one condition with unique singletons. This choice of stimulus displays may have reduced the influence of color redundancy on the probability of correct recall. Future studies should use a larger variety of patterns to investigate this issue.

Another reason to test a larger variety of patterns is that our metric could potentially account for the fact that color-similarity boosts can occur regardless of

distance between the repeated items (Morey et al., 2015; Morey, 2019). Algorithmic complexity could account for gradual effects depending on the distance between repeated items. Typically, based on an alphabet of four symbols, the algorithmic complexity for short strings predicts increasing complexities for blue-red-blue-green, blue-blue-red-green, blue-red-green-yellow, and blue-red-green-blue, respectively. Therefore, the alternating pattern blue-red-blue is estimated simpler than blue-blue-red when mixed with a third color (here, green), but a too large distance between the two repeated symbols like in blue-red-green-blue is estimated less compressible than blue-red-green-yellow. Future studies should therefore benefit from a larger sample of patterns which for instance include these types of patterns.

Conclusions

In summary, we examined whether the compression of information account could offer a general explanation of the color-sharing bonus in visual working memory, using objective algorithmic (Kolmogorov) complexity measures of visual displays. Our findings indicate that compression of information is a good candidate to account for our data, but an alternative explanation remains that the number of clusters could play a central role in the regularities that, we believe, algorithmic complexity can detect. Effectively, the two accounts offer quite similar fit of the data (one exception was that the compressibility factor resulted into a more precise explanation of the memory for four-item sub-patterns based on the complexity of the other four items). Either way, the compressibility factor and the number of clusters factor can both capture the effect of duplicates in visual displays on working memory optimization. This adds to the literature dedicated to understanding how improved memory may rely on specific computational processes aiming at detecting relational information on the spot. In sum, the superiority of the compressibility factor remains to be proven or else, more precise metrics should be developed in the future for two-dimensional displays to achieve better predictions. It also has to be noted that we only considered a lossless compressibility metric, but more complex lossy compression processes could be at play as well, as it seems to be the case for spatial locations (Haladjian & Mathy, 2015). This path might help increase the fit offer by the present compression account.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.

Barton, K., & Barton, M. K. (2015). Package MuMIn. *Version*, 1:18.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1–11.

Ben-Av, M. B., & Sagi, D. (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, *35*(6), 853–866.

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.

Brady, T. F., & Alvarez, G. A. (2011a). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011b). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*, 1–34.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*, 85–109.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science. *Trends in Cognitive Sciences*, *7*(1), 19–22.

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96–107.

Chekaf, M., Gauvrit, N., Guida, A., & Mathy, F. (2018). Compression in working memory and its relationship with fluid intelligence. *Cognitive Science*, *42*, 904–922.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Della Sala, S., Darling, S., & Logie, R. H. (2010). Items on the left are better remembered. *The Quarterly Journal of Experimental Psychology 63*(5), 848–855.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver oper-ating characteristic curves: a nonparametric approach. *Biometrics*, 837–845.

Dickinson, C. A., & Intraub, H. (2009). Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception, & Psychophysics*, *71*(6), 1251–1262.

Donkin, C., Tran, S. C., & Nosofsky, R. (2014). Landscaping analyses of the ROC predictions of discrete-slots and signal-detection models of visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 2103–2116.

Feldman, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica*, *102*, 137–163.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*, 227–232.

Gao, Z., Gao, Q., Tang, N., Shui, R., & Shen, M. (2016). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition*, *146*, 277–288.

Gauvrit, N., Singmann, H., Soler-Toscano, F., & Zenil, H. (2015). Algorithmic complexity for psychology: A user-friendly implementation of the coding theorem method. *Behavior Research Methods*, *48*(1), 314–329.

Haladjian, H. H., & Mathy, F. (2015). A snapshot is all it takes to encode object locations into spatial memory. *Vision Research*, *107*, 133–145.

Han, S., Humphreys, G. W., & Chen, L. (1999). Uniform connectedness and classical gestalt principles of perceptual grouping. *Perception & Psychophysics*, *61*(4), 661–674.

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *41*(2), 325–347.

Jeffreys, H. (1939). *Theory of probability*. OUP Oxford.

Jiang, Y., Chun, M. M., & Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics*, *66*(3), 446–453.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 683–702.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*, 1–13.

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*, 1–7.

Kubovy, M., & Van Den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological review*, *115*(1), 131–154.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, *51*, 57–74.

Li, M., & Vitányi, P. (2008). *An introduction to Kolmogorov complexity and its applications*, (Vol. 3). New York: Springer.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Mathy, F., Fartoukh, M., Gauvrit, N., & Guida, A. (2016). Developmental abilities to form chunks in immediate memory and its non-relationship to span development. *Frontiers in Psychology*, *7*:201.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*, 346–362.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Morey, C. C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology*, *3*(1), 18–29.

Morey, C. C. (2019). Perceptual grouping boosts visual working memory capacity and reduces effort during retention. *British Journal of Psychology*, *110*(2), 306–327.

Norris, D., Kalm, K., & Hall, J. (2020). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition* *46*(5), 872-893.

Norris, D. G., & Kalm, K. (2019). What's in a chunk? chunking and data compression in verbal short-term memory. *PsyArXiv. Advance online publication*.

Nuthmann, A., & Matthias, E. (2014). Time course of pseudoneglect in scene viewing. *Cortex*, *52*, 113–119.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition*, *41*, 1212–1227.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., . . ., Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750.

Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological review*, *120*(2), 297–328.

Peterson, D. J., & Berryhill, M. E. (2013). The gestalt principle of similarity benefits visual working memory. *Psychonomic bulletin & review*, *20*(6), 1282–1289.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, *26*, 303–343.

Quinlan, P. T., & Wilton, R. N. (1998). Grouping by proximity or similarity? competition between the gestalt principles in vision. *Perception*, *27*(4), 417–430.

Quinlan, P. T., & Cohen, D. J. (2012). Grouping and binding in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*(5), 1432–1438.

Reder, L., Liu, X., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review*, *23*, 271–277.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465–471.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12 :77.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*, 5975–5979.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–330.

Simmering, V. R., Miller, H. E., & Bohache, K. (2015). Different developmental trajectories across feature types support a dynamic field model of visual working memory development. *Attention, Perception & Psychophysics*, *77*, 1170–1188.

Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2013). Correspondence and independence of numerical evaluations of algorithmic information measures. *Computability*, *2*(2), 125–140.

Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2014). Calculating Kolmogorov complexity from the output frequency distributions of small Turing machines. *PloS One*, *9*(5):e96223.

Son, G., Oh, B.-I., Kang, M.-S., & Chong, S. C. (2020). Similarity-based clusters are representational units of visual working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *46*(1), 46–59.

van der Loo, M. (2016). stringdist: Approximate string matching and string distance functions [computer software manual].

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80–87.

Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, *64*(8), 1260–1280.

Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, *68*(5), 815–828.

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(47), 18766–18771.