Under review; please don't quote

Chunking of Categorizable Objects on the Fly

Mustapha Chekaf and Fabien Mathy
Université de Franche-Comté – France

This paper aims to evaluate the capacity of working memory while measuring the compressibility of information likely to allow the formation of chunks. It is hypothesized that the average working memory capacity (4 ± 1 chunks) allows the span to vary depending on the sum of information compressed within each chunk. We constructed a task of immediate serial report in which the participants had to memorize and recall sequences of categorizable stimuli, that is, objects with possible associations between them. The logical complexity of Feldman's (2000) rules of classification was used to measure the compressibility of each sequence of stimuli. To favor the extraction of informational regularities by participants, a second manipulation of the sequences concerned the presentation order. The results showed a span of about 4 objects, with a variability which depended on both the complexity of the sequences and their presentation order. The amount of information retained in working memory was maximal for the most compressible sequences and for the presentation orders that facilitated the abstraction of rules. The role of working memory in the formation of chunks is discussed.

It is accepted that individuals have a tendency to group information in order to make it easier to retain by recoding it in chunks (Anderson, Bothell, Lebiere, & Matessa, 1998; Cowan, Chen, & Rouder, 2004; Logan, 2004; Naveh-Benjamin, Cowan, Kilb, & Chen, 2007; Perlman, Pothos, Edwards, & Tzelgov, 2010; Perruchet & Pacteau, 1990; Tulving & Patkau, 1962). The process of chunking allows the simplification of a memorization task by taking advantage of regularities in order to reduce the quantity of information to retain (Ericsson, Chase, & Faloon, 1980; Miller, 1956). This process can for instance function via temporal segmentation (1.9.8.4 $\cong$ 19 − 84), reorganization (4.8.9.1 $\cong$ *reverse*(1984)), or pointing to preexisting knowledge in the long-term memory (1.9.8.4 $\cong$ Orwell's novel). As a key learning mechanism, chunking has had considerable impact on the study of expertise (Chase & Simon, 1973; Charness, 1979; Ericsson & Kintsch, 1995; Egan & Schwartz, 1979; Hu & Ericsson, 2012; Gobet et al., 2001; McKeithen, Reitman, Rueter, & Hirtle, 1981) and animal learning (Terrace, 1987, 2001). Much progress has also been made in the comprehension of temporal grouping on immediate recall (Farrell, 2008, 2012; Maybery, Parmentier, & Jones, 2002; Ng & Maybery, 2002). Grouping, instead of chunking, is a more neutral term that tends to indicate that no reference to long term memory is made. Simply, people tend to parse the to-be-recalled lists into clusters of temporally proximate information 'on the fly', which produces better recall of the clustered sequences that are mentally organized into mini-lists (which may be reflected by miniature serial position curves).

Little is known about the role of working memory in the *creation* of chunks, that is, the possibility that groups reflecting regular sets within a sequence can be used as germinal chunks 'on the fly'. Repeated exposure to the groups only favor the encoding of more permanent chunks in long term memory, but chunk creation is a mechanism that can be clearly separated from chunk retrieval (Guida, Gobet, Tardieu, & Nicolas, 2012). Whatever the origin of the recoding process (long term for chunks or short term for groups), the presence of regularity in the to-be-recalled material can account for apparent disparate capacities (this is the case for instance in visual working memory; Brady, Konkle, & Alvarez, 2009, 2011; Brady & Tenenbaum, 2013). One reason is that chunking (or grouping) has been controlled over the past 50 years in working memory tasks because it runs counter to a rigorous estimation of the mnemonic span (Cowan, 2001). The most basic way to control chunking consists of limiting the time of presentation and recall. Luck and Vogel (1997), for example, proposed a rapid task using image stimuli such as oriented lines or shapes of different colors. Their conclusion was that the amount of information that one can memorize in visual short-term memory does not depend on the number of features but is limited by four objects. Other works support the hypothesis that capacity is limited by the number of features (Magnussen, Greenlee, & Thomas, 1996; Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002). The present study focuses on determining what rules apply to grouping and how this relates to capacity.

Using a chunking memory span paradigm, Mathy and Feldman (2012) showed that capacity in working memory can be evaluated both in terms of chunks (each of the 4±1 chunks, Cowan, 2001, being fundamental units associated with slots in working memory, W. Zhang & Luck, 2008)

This research was supported by a grant from the Agence Nationale de la Recherche (Grant # ANR-09-JCJC-0131-01) awarded to Fabien Mathy. Correspondence concerning this article should be addressed to Fabien Mathy, 30-32 rue Mégevand, 25030 Besançon Cedex, France. or by e-mail to fabien.mathy@univ-fcomte.fr.

and in terms of the number of items that can be unpacked from the chunks (which can agree with Miller's estimation of 7±2 items in some circumstances, but not always), but that the real capacity is actually 4 (a result that concurs with certain theoretical analyses, Dirlam, 1972; Macgregor, 1988; Simon, 1974, and with other empirical studies, Broadbent, 1975; Gobet & Clarkson, 2004; Halford, Baker, McCredden, & Bain, 2005; Halford, Wilson, & Phillips, 1998; Luck & Vogel, 1997; Pylyshyn & Storm, 1988; Wickelgren, 1967; W. Zhang & Luck, 2008). The study by Mathy and Feldman (2012) encourages to see short-term memory and working memory as equivalent theoretical constructs, which is usually not the case (Aben, Stapert, & Blokland, 2012). Cowan's number refers to the number of chunks that can be stored after compression. Miller's number refers to the number of items before compression.

In the chunking memory span tasks developed by Mathy and Feldman (2012), chunking did not depend on the uncontrolled individual's perception of groups, but was rather induced by the regularity in the material. In this paradigm, the goal is to deliberately prompt the grouping of memory items, by allowing possible associations between items to form related units (for instance the list BBFFBBB could be retained as three chunks instead of seven independent letters). The idea is to introduce sequential patterns of variable length (e.g., 54321.246) to measure whether more regular sets can be represented more compactly by taking advantage of their regularity (see also, Mathy & Varré, 2013, Exp. 4, in which this paradigm is used to examine storage and processing together in association, with both processes dedicated to the to-be-recalled items). The present study also focuses on how capacity limitations can be overcome whenever a form of relational information can be computed in a sequence of items. However, instead of digits, the present study uses sequences of categorizable multi-dimensional artificial stimuli, with sequences offering possible associations between the stimuli. Following Mathy and Feldman (2012), the idea is that two non-independent items can be encoded in a more compact way into a single slot in memory, resulting in a greater number of available slots for the encoding of other items. The main characteristic of this paradigm is that the chunking process is quantified since regularities are known in advance.

The theory developed in this paper is that chunking processes are assessable by the compressibility of information (Mathy & Feldman, 2012). Three points must be clarified from the start about the notion of compression: 1) we aim to explain the formation of chunks on the fly (and not their use in relation to prior knowledge in long-term memory), 2) we describe a process of compression without loss of information (the exact original data can be reconstructed from the compressed data), and 3) we only attempt to quantify a process of logical compression. However, despite our use of a precise metric to describe how logical rules can be encoded, we do not lose sight of the fact that some verbal processes usually critical to short-term memorization (Baddeley, 1986; Burgess & Hitch, 1999; Baddeley, Thomson, & Buchanan, 1975; Chen & Cowan, 2005; Estes, 1973; Page & Norris, 1998; G. Zhang & Simon, 1985) are unfortunately left uncontrolled in our study. Still, we intentionally let the participants verbally recode our visual-based material because we thought that implementation of articulatory suppression could affect attention and subsequent encoding of the logical rules. It was our goal to let the participants encode the objects freely.

In the present study, two main factors are used to manipulate the complexity of the sequences. The first concerns their compressibility and the second the ordering of objects within sequences.

Based on the work of Feldman (2000) which provided a measure of conceptual complexity, we selected different categories of objects to be displayed and recalled serially. For instance: large white square, large black square, small white square, small black square, small white triangle, and small black triangle, which represents the category 'large squares or small objects'. We know the complexity of each of the category structures we used since Feldman developed a formalism of their underlying logical rule, which accurately predicts subjective complexity (i.e., the difficulty to learn the categories). Feldman showed that the subjective difficulty of a categorization task depends on the length of the shortest rule that describes the category, that is, the degree to which the category can be faithfully compressed. For example, using the same three binary dimensions (Size, Color, and Shape), the sequence of objects 'small black square, small black triangle, small white square, small white triangle' can be simplified by abstracting the feature common to the four objects: "small". If these four objects are positive examples of a category to be discovered, the rule "IF small THEN positive" allows one to separate these objects from negative objects (large white triangle, large black triangle, large white square and large black square). Hence, the information for this category is very compressible and does not require much mental effort to be retained. This means that learning the complete category of positive objects by rote is unnecessary since a simpler rule/definition ("small") allows one to encode the category. Conversely, less compressibility makes information difficult to summarize, so that incompressible categories do not allow any learning strategy based on an abstraction process to occur.

The second factor was order. Certain sequential presentations of the objects may hinder the abstraction of the category structure, when the objects do not seem to connect easily to form a given rule. As a result, the recall of the sequence seems more difficult. Effectively, although the compressibility metric provided by Feldman (2000) does not vary with the permutation of the stimuli of a given category, our hypothe-

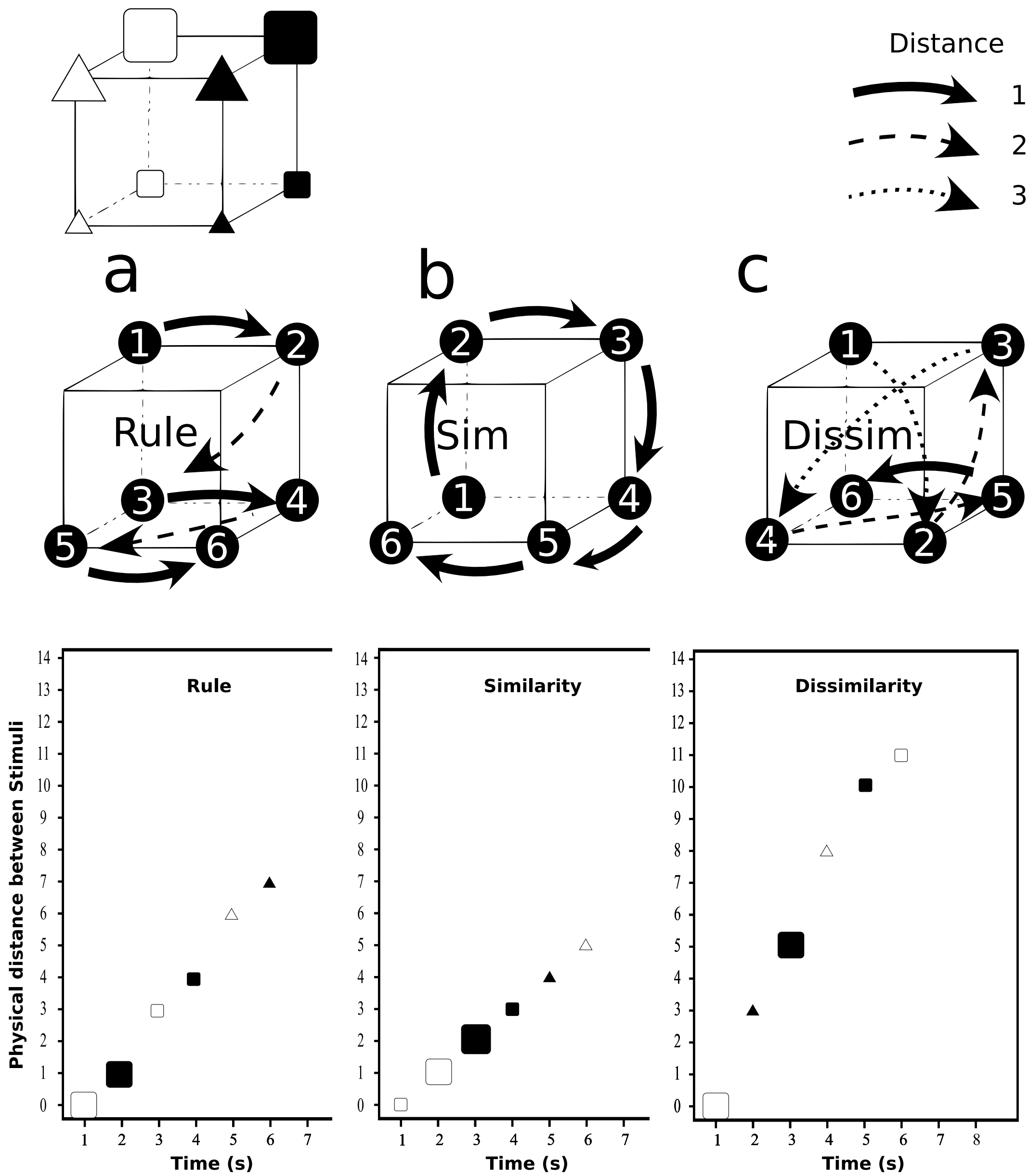

Figure 1. Example of construction of three sequences from the category structure “square or small”, representing six objects (top left cube). The category objects are replaced by black numbered circles on each of the three cubes under the (a), (b), (c) options, to simulate their sequencing. In each of the three cubes, one sequence (or presentation order) is indicated by the numbers from one to six as well as by the arrows. The distance between two successive objects is described by arrow type: continuous (an edge), broken (2 edges) and dotted (three edges). In the Rule condition, the objects are presented in three clusters within which the full arrows are parallel and in the same direction. This regularity is related to two factors: the separation of the clusters by discontinued arrows (2 edges) and the similarity between the objects of the same cluster, which both facilitate the formation of chunks. In the Similarity condition, the inter-stimuli distance is minimal. All the objects are linked between them by continuous arrows. In the Dissimilarity condition, the sequence is characterized by a maximal inter-stimuli distance. The graphs show the distances separating the stimuli as a function of presentation time (1 second in the condition). These graphs show the intermediary aspect of the mean inter-stimulus distance in the Rule condition. In the Rule condition, distance leaps can facilitate the identification of clusters. The intervals are more numerous and less regular in the Dissimilarity condition.

sis is that individuals only benefit from compressibility when the presentation order follows the process of simplifying the information that they attempt to make use of. The second idea was therefore to order the stimuli of a given category according to three possible options: rule-based, similarity-based (i.e., step by step), and dissimilarity-based (i.e., disorganized). Mathy and Feldman (2009) showed that learning categories was more effective when information was presented in a rule-based fashion, in comparison to a less effective order maximizing inter-item similarity (Elio & Anderson, 1981, 1984; Gagné, 1950; Medin & Bettger, 1994), and worse in a dissimilarity-based fashion. In making the hypothesis that participants use logical rules in order to reduce information, we predict that an order facilitating the formation of a logical rule will favor the chunking of objects and consequently their recall. The main difference between research on categorization and the present study is that instead of being required to recall an unordered set of objects belonging to one of two contrasting categories, our participants were required to memorize a sequence of objects of a given category. According to our hypothesis, it is easier to unpack the stimuli from a well-structured rule. For instance, "square[small:...[black:...]]" may be a compressed hierarchical description for retaining the set "small black square, small white square, large black square, large white square" (the : symbol for instance indicates that there are other values to be listed after the one that precedes it). The "small:...[black:...]" part simply means that the square objects have to be listed beginning with all the small ones (then the non-small ones), and beginning with the black one for every size (then the non-black one). The only way to recover the original information is to follow the correct algorithm to unpack the sequence. Ideally, this works for individuals: for instance, if one plans on cleaning two bedrooms, beginning by straightening up then vacuuming, starting by the smaller room (bedroom #1). Then, the following order can be executed: straightening up bedroom #1, straightening up bedroom #2, vacuuming bedroom #1, vacuuming bedroom #2 (by analogy: "clean[straighten up:...[bedroom #1:...]]"). The present study tends to show that participants are able to encode and decode information hierarchically according to this expectation in short-term time.

In sum, we expect that recall will be more effective for compressible sequences (Feldman, 2000) and particularly in the rule-based presentation order condition (Mathy & Feldman, 2009), while agreeing with the prediction of a mean recall that fits the 4 ± 1 capacity (Cowan, 2001, 2010). Our results will show that it is conceivable to measure short-term memory capacity and chunking simultaneously.

Experiment 1

The computerized experiment included 51 sequences with one to eight stimuli presented serially. Each sequence was followed by a recall phase in which the stimuli had to be recalled serially. The stimuli were categorizable multi-dimensional objects with discrete features (e.g., large black triangles, etc.). These stimuli were of particular interest as their categorization has been investigated quite thoroughly for more than half a century (Bruner, Goodnow, & Austin, 1956; Shepard, Hovland, & Jenkins, 1961). For instance, there is consensus that the predicted complexity of different categories for 3-dimensional objects globally fits the difficulty the participants have in learning the categories. This domain therefore provides a precise compressibility metric, as opposed to lists of words for instance. The two main factors in the present experiment were the complexity structure of the 3-dimensional categories and the presentation order within categories. Another basic factor was the number of objects per category.

Participants. Sixty-seven students enrolled at the University of Besançon, $M = 22$ years old ($sd = 2.7$), volunteered to participate in this experiment.

Stimuli. Our stimuli could vary according to three dimensions: shape, size and color. There were only two possible sizes (280×280 pixels or 140×140 pixels). For a given sequence, the program randomly chose two out of eight shapes and two out of eight colors (cf. Fig. 2.a), in order to create a set of eight objects. For example, if the values triangles, squares, white and black were drawn, the program generated $2 \times 2 \times 2 = 8$ stimuli by combining three features for each stimulus (small white triangle, large white triangle, ..., large black square). These dimension values made it possible to generate 1568 possible sets of eight objects, so that the probability of a participant meeting two identical sets during the entire experiment would be very low. The stimuli were presented on a gray background.

Categories. Truncated categories from 1 to 8 stimuli were drawn from the initial set of 8 stimuli. The categories were then organized and transformed into sequences according to a procedure explained below. In short, there was a threefold distinction for each sequence: a set of stimuli (i.e., the original set made of eight stimuli), a category (a subset of the original set of eight stimuli) and a sequence (a subset in which the stimuli were ordered).

Any set of eight 3-D stimuli can be represented schematically on a cube called a Hasse diagram (cf. Fig. 1), whose three dimensions (height, length and depth) can be associated with those of our objects (size, color and shape). If two objects are situated on two adjacent vertices, then they only differ by a single dimension. For example, a large black square and a large black triangle only differ by their shape. The shortest distance between two objects (by following the edges) is equivalent to the number of differences which separate them. A large black triangle and a small white square,

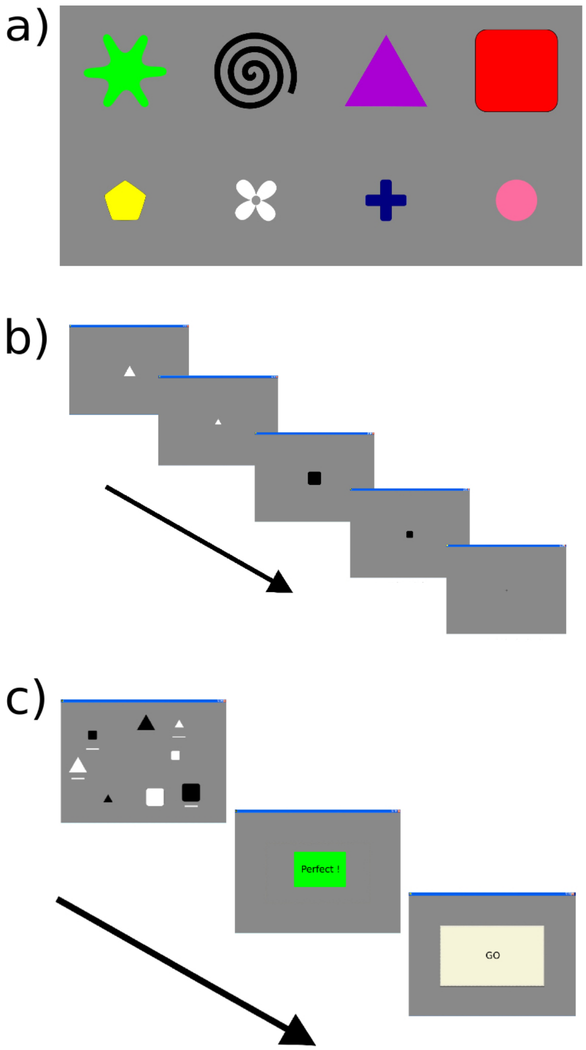

Figure 2. (a) A sample of the 8 types of shapes, 8 types of colors and 2 types of sizes on which the stimuli were constructed. (b) An example of a sequence of 4 stimuli, followed by a screen displaying a fixation cross. (c) Screen of 8 randomly placed stimuli including the 4 stimuli presented during the preceding to-be-remembered sequence; the stimuli are underlined using a white bar when the user clicks on them; when the user validates his answer with the space bar, a feedback screen informs the participant if the recall is correct (here: Perfect ! In green); a last screen with a GO window waits until the user moves on to the next sequence by pressing on the space bar.

for example, differ by their shape, size and color, and they are physically separated by three edges on the diagram. In the Hasse diagram, the positive examples are represented by the vertices marked by a black circle and the negative ones are represented by the empty vertices. The Hasse diagram is therefore very useful to represent the sparsity of objects within categories. This sparsity directly impacts the category complexity.

The to-be-recalled categories of stimuli were chosen on the basis of the exhaustive list of the 21 categories shown in Figure 3. Figure 3 also gives Feldman's (2000) measure of compressibility (FC) for each category, which was used as a measure of complexity. FC is indicated in the middle column because the FC numbers for a given line fit both the category structures on the left and those on the right. Note that FC is not confounded with sequence length for sequences equal to 2, 3, 4, 5 and 6 objects. However, for the categories made of 1, 7 or 8 objects, no variations of complexity could be manipulated. When a set was composed of eight objects, categorization was no longer involved. In this particular case, the shortest formula was trivial (i.e., "all objects") and it was associated with a not applicable (N/A) FC value. Still, retaining a sequence of 8 objects in our experiment required encoding their position: the complete set was presented at test when we asked participants to reconstruct the order of presentation.

Ordering of the categories in sequences. Again, a categorization task is akin to a free recall task: the participants are usually required to freely recall the objects of one category. In the serial recall task used in the present study, recalling several items in the correct order was more complicated than making a category. The measure of complexity only serves here to predict the chunkability of the category, not the memory of the sequence.

A second manipulation of the experiment consisted of sequencing the objects according to three conditions: Rule, Similarity and Dissimilarity. Before presenting these types of orders in greater detail, it is important to recall the interest of this manipulation. The compressibility of information is theoretically independent from the presentation order. However, controlled presentation orders were used (as opposed to using a random order) in order to manipulate the learnability of the categories, in other words, the participants' ability to make groupings of information, which is necessary for simplifying the information. According to the hypothesis, certain types of order favor the formulation of a rule describing the objects presented, and, as a consequence, their recall. The 21 categories of objects were transformed into 51 sequences according to these three presentation orders. The number 51 results from the fact that the order of some categories (six categories in all) was not manipulable. This was the case, for example, with categories made up of one or two

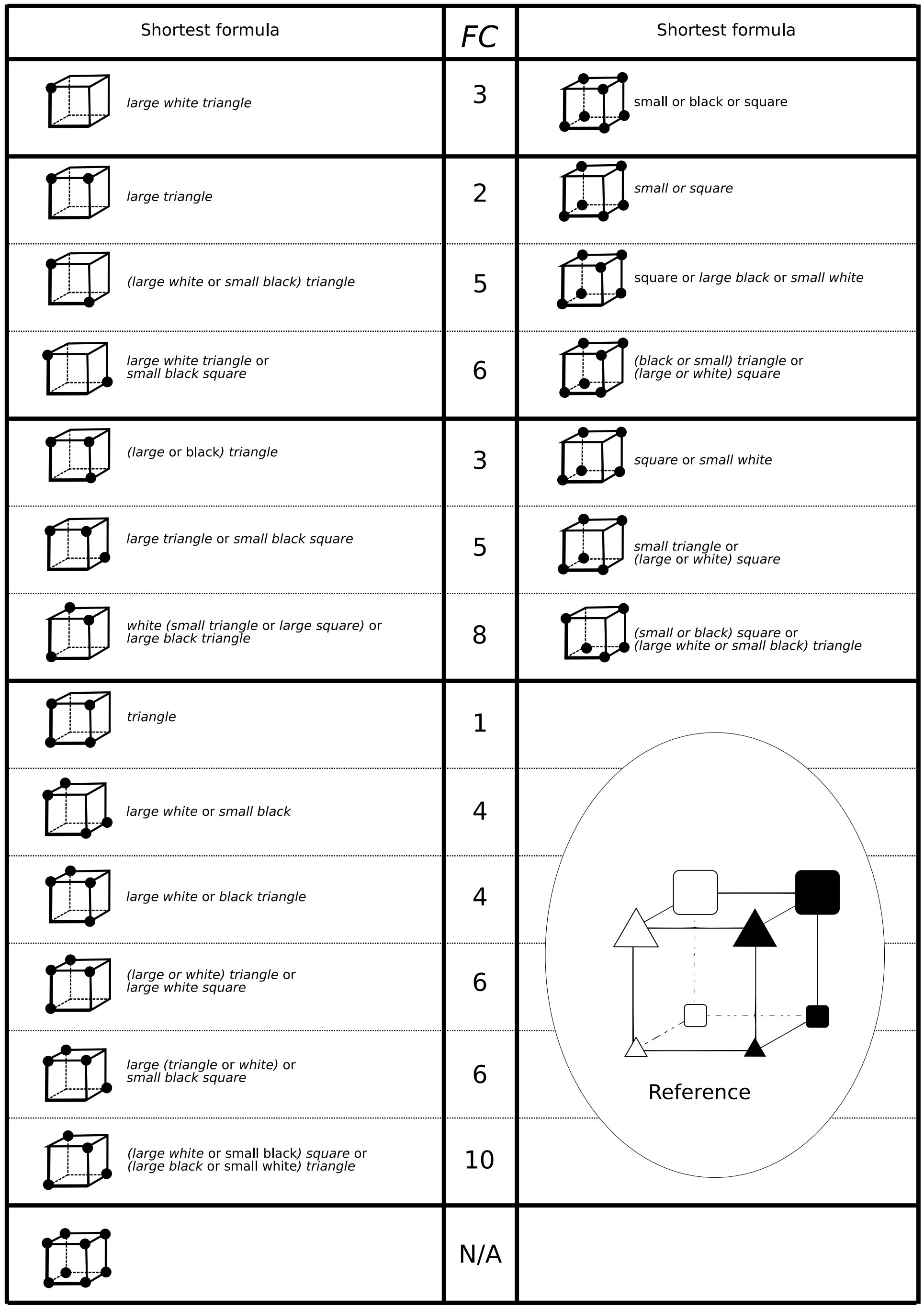

Figure 3. Column FC indicates Feldman's (2000) measure of logical compressibility for each of the categories in the left or right column. This measure only indicates the number of literals in the shortest logical formula allowing one to describe the positive examples of a category indicated by a black circle. The black circles positioned on the cube illustrate the relationship between the objects that could be used by the participants to associate the stimuli in our experiment. Each of the stimuli selected for a category were presented sequentially and their location remained constant in the middle of the screen, so no reference was made during the experiment to the spatial organization shown in this figure (this was done to show the similarity structure between objects more easily). For instance, the 8th cell on the left indicates that four triangles were used to construct the memory list (cf. the bottom right cube above 'Reference' for retrieving which of the stimuli the black circles in the 8th cell on the left point at). The "triangle" category representation would be sufficient for a free recall of the list. However, the observers were asked to reconstruct the sequence, which required a more precise encoding of the list.

objects (once the first object is drawn, only the second object remains), and also of categories for which the distances are identical between any pair of objects (this is the case for the category in the left column of Fig. 3 for which FC = 8 and FC = 10). These six categories were directly converted into "order condition = None" coded sequences. The Rule, Similarity and Dissimilarity conditions were applied to the 15 remaining categories. Therefore, each of the participants were only administered 51 sequences ($6 + 3 \times 15$).

The three presentation orders that were based on a study by Mathy and Feldman (2009) are detailed below. For all presentation orders, note that the similarity between two objects can be evaluated by the number of edges which separates them on a Hasse diagram (i.e., a city-block distance).

In the **Rule condition** (Fig. 1), the objects were grouped by clusters (i.e., by sub-categories). The similarity between objects within clusters was maximal while it was minimal from one cluster to another. It was hypothesized that a minimal similarity between clusters contributes to greater discrimination of the clusters into distinguishable units, thus allowing the formulation of a rule. Figure 1 shows the example of an organized sequence in the Rule condition. The shortest rule for describing the set of objects to be freely recalled (i.e., not serially) is "square or small". In the Rule condition, the objects in this sequence were presented in an order which organized them successively into pairs corresponding to sub-clusters. Each pair presented two objects which differed only by their color, first the white object, then the black one. The squares were presented first (the pair of large squares, then the pair of small squares, each pair representing a sub-cluster of the cluster "square"), followed by the small triangles (a sub-cluster of the cluster "triangle"). Therefore, the differences were minimal within sub-clusters (a single difference in color), but more marked between sub-clusters. For example, a small black square followed by a small white triangle (two feature differences) marks a separation between two clusters. The distance between the large black square object and the small white square was also made of two features in this example. In the Rule condition, distance leaps (cf. Fig. 1, bottom) can facilitate the identification of clusters and consecutively it might induce the formation of chunks. The Rule condition urged the participants to reorganize the "square or small" rule into a more hierarchical rule made up of a "white, black" iteration embedded in a "large squares, small squares, small triangles" iteration. By looping all the elements of both arrays (totalizing 8 features), the participant could recover the ordered category of the original six objects comprising 18 features. With a bit of practice, the rememberer could further increase the reduction of information by adopting a more personal formulation such as "(white,black) × [large squares + small (squares + triangles)]" that reduced the number of features to seven. We do not assume that the participants are using this exact type of encoding, but rather we indicate that any regularity could be encoded by the participants. Another less compressible material made of six other objects, e.g., (black or small) triangles or (large or white squares)" ordered in a nonregular way ("large black square, small white triangle, small black triangle, large white square, large black triangle, and small white square") was definitely non-chunkable. We hypothesized that the Rule condition presents successive objects in organized, distinguishable and thus chunkable clusters. In this condition, according to our expectations, the participants should have the best recall performance, particularly if they attempt to formulate a rule for memorizing the objects[1].

In the **Similarity condition**, the first object of the sequence was chosen to favor a minimal inter-stimuli total distance, each object that followed presented a maximal similarity with the preceding one, so that the global distance uniting all the stimuli was the shortest possible. The objects were then presented in a chain following the principle that there was maximal similarity between two successively presented objects. Following Mathy and Feldman (2009), although the mnemonic traces of the objects within sequences may be implicitly reinforced by their temporal contiguity in the Similarity condition, resulting in the memorization of exemplars (Medin & Schaffer, 1978; Nosofsky, 1986), it is hypothesized that the participants should show poorer results than in the Rule condition, particularly because the participants are inclined to form explicit verbal rules rather than trying to retain exemplars in such categorization tasks (Ashby & Ell, 2001). Figure 1 shows an example of Similarity order for the same category of objects. In this example, the objects are presented successively with a single feature difference between each stimulus. Note that because there are several ways (24) in which the stimuli can be ordered by maximizing the similarity between the four stimuli in the cube, there is no way of retaining a way to encode a path for the eight chained stimuli.

Finally, in the **Dissimilarity condition**, the first object was chosen so as to favor a maximal inter-stimuli total dis-

[1] This rule-based presentation order is a bit different from the one used by Mathy and Feldman (2009). Because the study by Mathy and Feldman (2009) focused on categorization processes, the ordering of the objects within clusters was not ordered in such a logical fashion, but randomly. Mathy and Feldman (2009) hypothesized that the randomization of that which is not informative to form a category (steps within a cluster) yields superior learning. For instance, in an induction task, if a red bird is followed by a red car, the abstraction "red things" might form more readily than a if a red bird is followed by a red fish which tends to orient the learner toward a more specific concept such as "red animals". Because the present tasks required the retention of the correct order of the clusters and the correct order of objects within clusters, our choice consisted in segregating the clusters while ordering logically the objects within clusters to urge participants to form a hierachical rule.

tance. Each object that followed presented a minimal similarity with the preceding one, so that the global distance between stimuli was maximal. This principle of presentation can be considered as voluntarily disorganizing the presentation so that the associations between stimuli are more difficult. This condition should perturb the memorization of objects. Consequently, performance in terms of recall should be lower than in the other two conditions. Figure 1 shows an ordered sequence in the dissimilarity condition using the same example as above.

Note that the manipulation of the presentation orders was less effective (if possible) for the sequences presenting the highest FC simply because the higher heterogeneity of the categories hindered their manipulation. For instance, the white (small triangle or large square) or large black triangle concept (7^{th} cell in the left column of Figure 3) was typical of a concept for which different presentation orders could not be specified. Figure 4 illustrates this nonindependence between compressibility and order. The most incompressible categories (top of the triangle) provides fewer possibilities of order manipulation. The reason is that once a first object is chosen, there is no other choice than picking a second object that is two-feature away from the first one, and so on. The top category in Figure 4 indicates therefore a $R = S = D$ configuration that illustrates the idea that rule-based, similarity-based and dissimilarity-based orders lead to the same unique possible ordering type. The second row in Figure 4 states that a dissimilarity-based order is qualitatively different from the two other orders (rule-based and similarity-based). The rule-based order is almost identical to the similarity-based order, as only one permutation makes a difference: the rule-based order is "front-top-left, front-bottom-left, front-top-right, and back-top-left", whereas the similarity-based order is "front-bottom-left, front-top-left, front-top-right, and back-top-left". The only difference is therefore 'front-bottom-left, front-top-left' vs 'front-top-left, front-bottom-left' for the first two objects, making both presentation orders almost identical (which we represented using a $R \simeq S$ notation). The last row, where FC = 1, allows more diverse manipulation of orders, resulting in a clear demarcation between the three types of orders.

Procedure. Serial report was investigated using a procedure similar to the visual STM serial report task (Avons & Mason, 1999; Smyth, Hay, Hitch, & Horton, 2005). The computerized session included 51 sequences with one to eight stimuli presented serially at intervals of one second (26 participants) or two seconds (41 participants) per stimulus depending on the condition[2]. Before beginning, the participants were asked to read a description of the experiment which informed them that sequences of images would appear on the screen, that they had to memorize the images in the correct order, and that each phase of presentation would be immediately followed by a recall phase.

In the phase in which the sequence was presented, the stimuli were displayed in the center of the screen (cf. Fig. 2.b) for one or two seconds. Then a fixation cross appeared on the screen for one second. This phase was followed by a recall phase (cf. Fig. 2.c) during which the original set of eight stimuli was randomly placed on the screen. The stimuli were underlined when the user clicked on them. After the user validated his answer with the space bar, a feedback screen indicated if the recall was correct (i.e., item memory + order memory both correct), then a screen with a GO window appeared and the user moved on to the following sequence by pressing on the space bar. The order and the response times were recorded automatically by the program.

The 51 sequences were presented in random order to avoid any ascending presentation of the length or the complexity of the items. Moreover, concerning the distribution of the items on the Hasse diagram, we established six possible rotations for each category structure that corresponded to six possible ways to place the objects on the diagram while preserving the same structure. A rotation was randomly drawn for each of the sequences presented in order to multiply the possible combinations in dimensional terms (shapes, sizes and colors). Thus, whatever the sequence, the participant did not know in advance which of the dimensions would be the most pertinent for recall. The experiment lasted on average 25 minutes. In all, 3417 sequences/trials (51 $\times$ 67 participants) were presented.

Results

The following analyses were conducted on serial recall scores coded correct (1) or incorrect (0) for each trial (a response was scored correct when both the items and the positions were correctly recalled), proportion correct (when the recall scores were averaged across conditions) and response times (the entire amount of time needed to recall a sequence). In the following Results section, we sometimes kept all the trials in the analyzes in order to determine the whole data variance, to get more power, or when our goal was to refine the basic ANOVA that were run on aggregated data. When we used within-subjects or beween-subjects ANOVA, we first aggregated the data for a given variable of interest, for instance 'presentation time', to run separate univariate ANOVA for each dependent variable of interest (e.g., the mean percent across trials, or the mean RT).

[2] The one-second condition tends to be standard to measure the memory span, so we used this condition with a great number of participants. However, we still wanted to modulate the time allotted to chunking by extending the presentation rate (i.e., 2 s) with a smaller group of participants, because presentation rate can affect memorization strategies (Hockey, 1973)

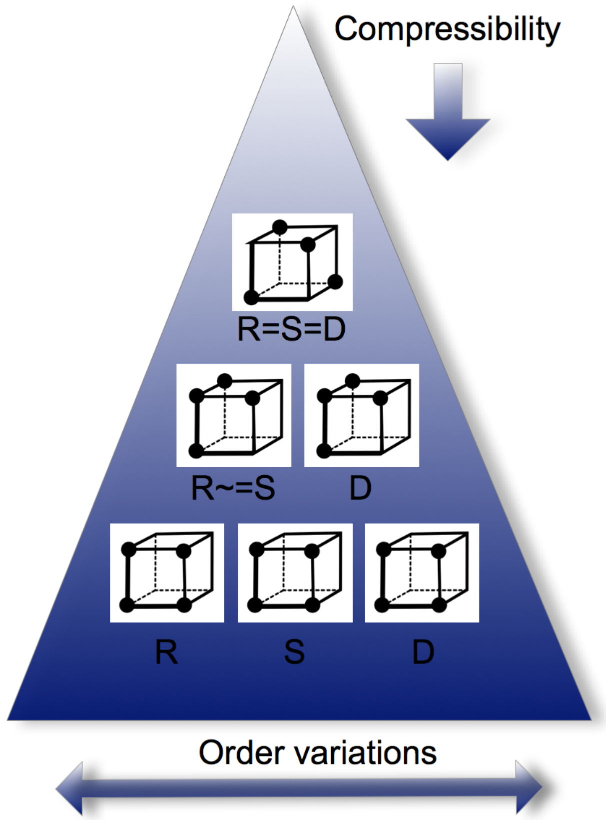

Figure 4. Manipulation of presentation orders is facilitated by more compressible categories. The $\sim =$ notation intends to represent the $\simeq$ notation.

Summary of the expected results: 1) Recall performance primarily depends on sequence length, because little information about order can be compressed, 2) A higher compressibility of information (i.e., lower FC within sequences of the same length) allows better recall because lower FC allows better recoding of the entire set of items, 3) More regular presentation orders (i.e., rule-based, followed by similarity-based and dissimilarity-based) favor the compression of the available regularities and simplify the encoding of item position.

Proportion correct. Analysis on proportion correct (based on the trial-by-trial 0/1 scores averaged across trials) was significant ($t(65) = -2.5, p = .014$), with respective means equal to .33 ($sd = .11$) and .27 ($sd = .08$) for the 2s and 1s conditions, with a size effect inferior to 10%, $\eta^2 = .09$.

Figure 5 shows the mean proportion of correct responses as a function of the length of the sequence presented, separated by the type of presentation order. Overall, a nonlinear regression (using an s-shape sigmoid function of the form $a - 1/(1 + exp(-b \times x + c))$ on the mean points of this figure (along the x axis) showed that 78% of the variance of performance was explained by Length ($R^2 = .78$). Independent of the orders, the estimation of the mnemonic span at a threshold of 50% of correct responses was 3.5 objects, which corresponds to the limit of 4 ± 1 chunks estimated by Cowan (2001) and to a previous estimation of the digit span resulting from a similar chunking memory span paradigm (Mathy & Feldman, 2012). When the length of the sequence increased from three to four objects, the mean performance dropped from .64 to .33. It also diminished by half again with between four and five items to recall (from .33 to .14). Thus, the mean performance was divided by 4 (from .64 to .14) when the length increased from three to five. This impressive drop of performance at list length = 4 recalls Cowan's (2001) estimation of capacity.

A multiple linear regression analysis on mean proportion correct showed that each of the three factors contributed significantly to the drop in performance. The respective percentages of variance explained were: length of the sequence ($\beta_i r_{Yi} = 70\%$), compressibility/FC ($\beta_i r_{Yi} = 9\%$, which included eight unique values, that is 1, 2, 3, 4, 5, 6, 8, 10), and presentation order[3] ($\beta_i r_{Yi}$ = 11%), totalizing 90%, $F(5, 44) = 72, p < .001, R^2 = .90$. The effects of length and presentation order on performance are clearly visible in Figure 5, whereas the effect of compressibility is shown in Figure 6. Regarding the effect of FC, a partial correlation analysis showed that at a constant sequence length, compressibility and correct proportion were negatively linked ($r = -.28, p < .001$). Effectively, the more the complexity of the sequence increased, the more the performance diminished (see Fig.6).

The ANOVA for repeated measures with FC as within-subject factor and proportion correct as dependent variable was significant for most sequence lengths: $F(2, 132) = 2.9, p = .054, \eta^2 = .04$, for two-object sequences; $F(2, 132) = 11.5, p < .001, \eta^2 = .15$, for three-object sequences;$F(3, 198) = 45.6, p < .001, \eta^2 = .41$, for four-object sequences;$F(2, 132) = 18.8, p < .001, \eta^2 = .22$, for five-object sequences; $F(2, 132) = 45.1, p < .001, \eta^2 = .41$, for six-object sequences. The mean data points for the latter analysis are reported in Figure 6. Fig. 6 (bottom right) shows the linear trend that we obtained using aggregated data across participants, complexity and list lengths, totalizing 1273 data points, ($F(2, 1270) = 893$, $p < .001$, $R^2 = .58$; $\beta_{FC} = -.33$; $\beta_{Length} = -.76$), both betas being significant.

The proportion correct for the Rule, Similarity, and Dissimilarity orders was respectively .35 (sd = .16), .25 (sd = .13), .16 (sd = .09). The ANOVA for repeated measures was significant, $F(2, 132) = 80, p < .001, \eta_p^2 = .55$. We also observed significant differences between the Rule and Similarity conditions, $t(66) = 6.8, p < .001$, and between the Similarity and Dissimilarity conditions, $t(66) = 7.8, p < .001$, which were still significant after considering the Bonferroni correction.

At lengths 3 and 4 (the conditions for which the four types of order applied), the reason for the lower score for the None order is linked to the complexity of the heterogenous categories for which FC = 8 and FC = 10. In these conditions, the proportion correct for the Rule, Similarity, Dissimilarity, and None orders was respectively .52 (sd = .03), .47 (sd = .03), .32 (sd = .02), and .28 (sd = .04). The ANOVA for repeated measures was significant, $F(3, 198) = 25, p < .001, \eta_p^2 = .28$. We also observed significant differences between the Rule and Similarity conditions, $t(66) = 2.11, p = .04$, between the Similarity and Dissimilarity conditions, $t(66) = 6.7, p < .001$, but the difference between the Dissimilarity and None conditions was not significant. At lengths 3 and 4, the only pair left significant after the Bonferroni correction was Similarity-Dissimilarity.

Response times. The mean response time for a whole sequence was 9.79 s. Overall, response times were higher in the 2s condition, ($t(65) = -2.4, p = .020$), with respective means for the 2s and 1s conditions equal to 10373 (sd = 2584) and 8864 (sd = 1985), with a size effect inferior to 10% ($\eta^2 = .09$).

The mean response times as a function of the Rule, Similarity and Dissimilarity orders were respectively: 9797 ms (sd = 2561), 9979 ms (sd = 2611), and 11439 ms (sd = 3453). The ANOVA for repeated measures applied to this data was significant, $F(2, 132) = 27, p < .001, \eta_p^2 = .29$. Paired tests between the order conditions showed nonsignificant difference in response times between the Rule and Similarity conditions, but a significant difference between the Similarity and Dissimilarity conditions, $t(66) = -5.5, p < .001$.

Likewise, Feldman's complexity significantly influenced the response times, $F(8, 528) = 25, p < .001, \eta_p^2 = .28$ (the ANOVA was based on all the FC values including the N/A one). When the complexity increased, it took the participants longer to recall the items. When presentation order and FC were both taken into account in the ANOVA, we also found a significant interaction, $F(14, 1056) = 7.7, p < .001, \eta_p^2 = .09$ that tended to show that the Rule-order benefited the participants more when FC was low. For instance, proportion correct was .49 higher for Rule-based presentation orders when FC = 2 (in comparison to the average of the similarity- and dissimilarity-based orders), whereas it was only .008 higher when FC = 8. The differences in proportion correct were .23, .29, .49, .14, .19, .05, .08, and .008 for the respective FC values: N/A (when 8 objects had to be recalled), 1, 2, 3, 4, 5, 6, and 8.

Position curves. Serial position curves are often scrutinized to explore serial recall, especially when grouping is involved (Anderson & Matessa, 1997; Frankish, 1985; Hitch, Burgess, Towse, & Culpin, 1996; Henson, Norris, Page, & Baddeley, 1996; Ng & Maybery, 2002; Maybery et al., 2002; Ryan, 1969). Hence, we now provide an analysis of serial-position curves, based on the idea that a chunking process can result in unusual serial-position curves. We did not expect large distorsion of the usual position curves such as multiple bowing with primacy/recency effects within each group that are sometimes observed when a temporal grouping is imposed on the items (for instance, mini serial position curves can be observed within groups of three items). One reason is that our use of Boolean dimensions tends to induce the grouping of objects by pairs or quadruplets only. A simple analysis can therefore search for runs of two or four items in order to count the chunks that have correctly been formed. To keep this analysis as simple as possible, we first focused on two-object groups to assess the strength of associations between two contiguous objects (Cowan et al., 2004; Chen & Cowan, 2005). We assessed the effect of the manipulation

[3] Because the use of dummy variables is more appropriate to represent multiple groups (Cohen & Cohen, 1983), the Rule, Similarity, Dissimilarity and None conditions were respectively recoded using three dummy variables.

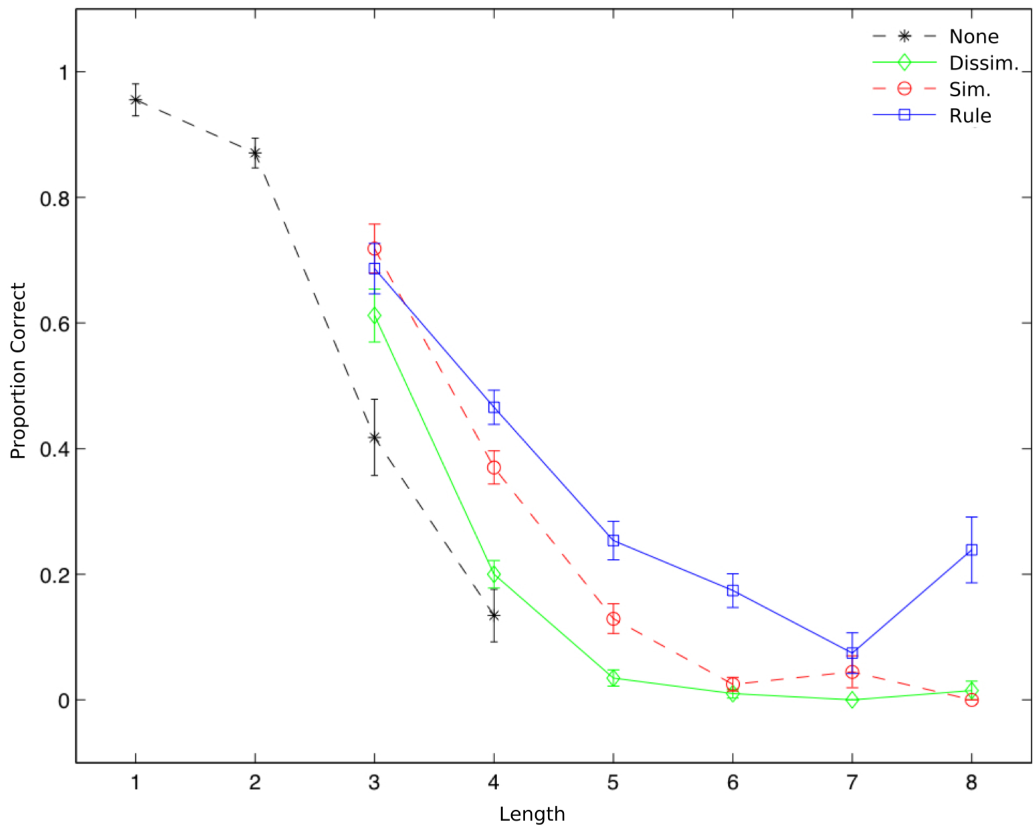

Figure 5. Proportion of correct sequences recalled as a function of the length of a sequence. The error bars show ±1 standard error.

of presentation order on list recall using strict serial-position scoring (i.e., an object was counted as correctly recalled only if in the correct position). For list lengths equal to 4, there were two potential two-object chunks (at positions 1-2 and 3-4); for list lengths equal to 6, there were three potential two-object chunks (1-2, 3-4, and 5-6); for odd list lengths, the last item was left out from the statistical analysis (although we kept the descriptive statistics in the plots). We tested whether there was a greater chance for participants to correctly recall in order the adjacent objects within these pairs. Our prediction was that the incentive to form chunks was a monotonic function of the dissimilarity-, similarity- and rule-based orders.

Figure 7 shows the serial-position curves (mean proportion correct) as a function of list length and presentation order. The blue, red and green colors stand for Rule-based, Similarity-based, and Dissimilarity-based respectively, to follow Fig. 5. The plots are supposed to be more stair-like (by pairs of positions) in the rule-based order condition. To verify this prediction, we counted the number of times that the pairs were correctly combined and recalled in order, across all possible pairs, as a function of presentation order. Figure 8 shows the proportion of perfectly recalled pairs (colored bars). A vertical comparison of the plots indicates that the proportion is progressively higher for the similarity- and rule-based presentations. When summing across list length, this analysis indicates that 840 pairs were chunked in the rule-based presentation order (whereas, 902 pairs were not chunked); 626 pairs in the similarity-based presentation order (against 1116); 365 pairs in the dissimilarity-based presentation order (against 1377). This 2×3 crosstab of the number of chunked pairs vs the number of nonchunked pairs accross order types yielded a significant statistical nonindependance between order type and chunking ($\chi^2(2) = 285$, $p < .001$; gamma ordinal test, .39, $p < .001$) showing that there was an advantage for recalling pairs of objects for similarity-based and rule-based presentation order. The odds-ratio between the rule-based order and the dissimilarity-based order indicates a basic size effect: chunking a pair was $(840/902)/(365/1377) = 3.5$ times as likely in the rule-based condition than in the dissimilarity condition.

A similar pattern of results was observed when we used a less strict scoring procedure, by counting one pair as chun-

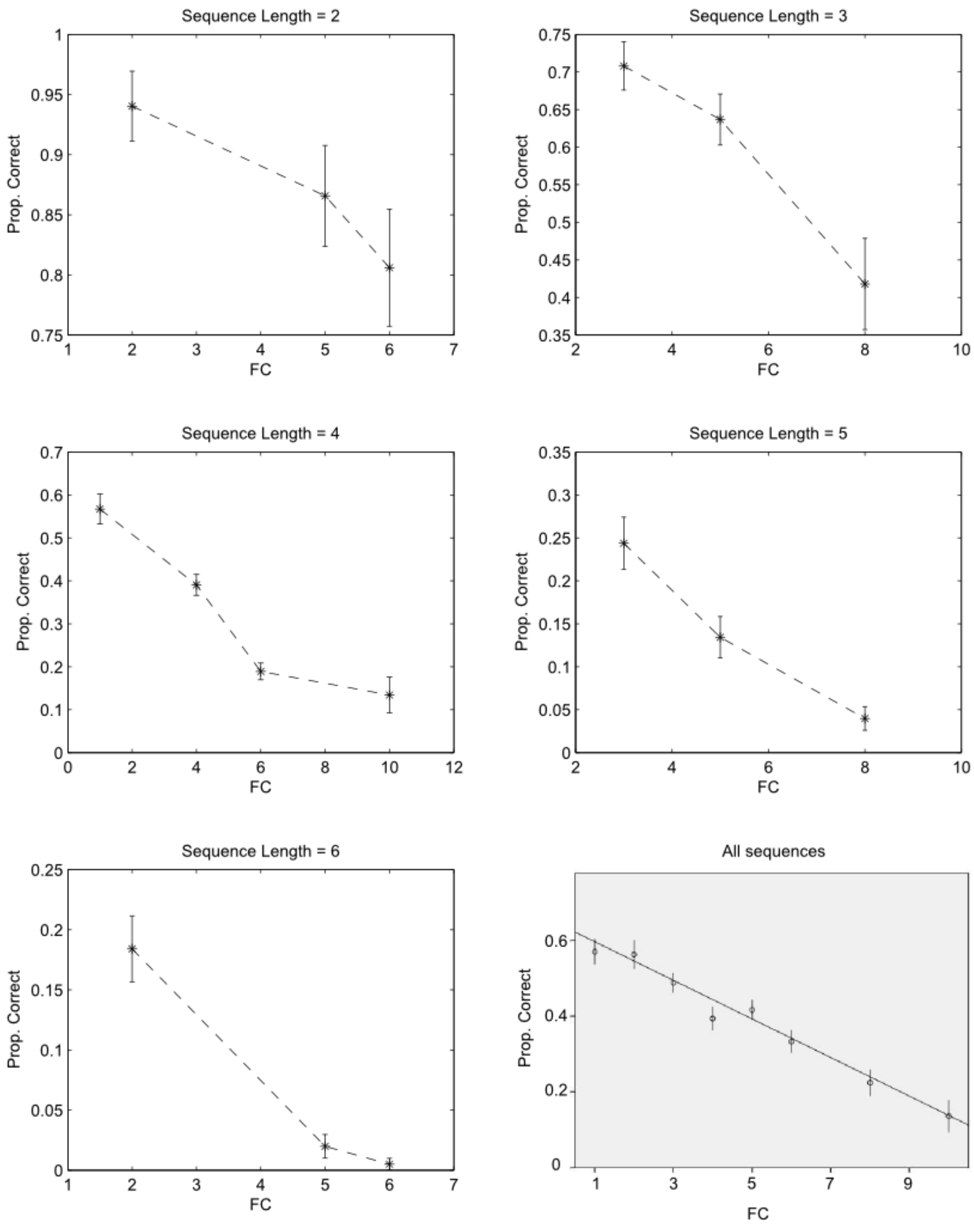

Figure 6. Proportion of correct sequences recalled as a function of Feldman Complexity (FC), at constant sequence lengths. Means and standard errors are based on the 3217 trials to include the whole variance. The error bars show ± one standard error.

ked even when it was not recalled at the correct position. For instance, when a 1-2 pair in the list was recalled in correct order within the pair, but at incorrect positions, for instance 2-3, the chunked was counted as correctly recalled. In this case 972 pairs were chunked in the rule-based presentation order (whereas, 770 pairs were not chunked); 696 pairs in the similarity-based presentation order (against 1046); 451 pairs in the dissimilarity-based presentation order (against 1291). This 2×3 crosstab also yielded significant statistics ($\chi^2(2) = 324$, $p < .001$; gamma ordinal test, .40, $p < .001$) showing a same advantage for more regular orders.

We then focused on the first quadruplet (items on positions 1-4). Figure 9 shows the proportion of perfectly recalled quadruplets at positions 1-2-3-4. The proportion is again progressively higher for the similarity- and rule-based presentations at constant list lengths. The crosstab indicated that the quadruplets were chunked 290 times in the rule-based presentation order (whereas, 380 quadruplets were not chunked); 182 times in the similarity-based presentation order (against 488); 101 times in the dissimilarity-based presentation order (against 569). The 2×3 crosstab of the number of chunked quadruplets vs the number of nonchunked quadruplets showed similar effects ($\chi^2(2) = 132$, $p < .001$; gamma ordinal test, .45, $p < .001$). The same odds-ratio indicated a $(290/380)/(101/569) = 4.3$ times more chance of chunking the first quadruplet in the rule-based condition than in the dissimilarity condition.

Experiment 2

The aim of this experiment was to adapt Experiment 1 using a different procedure for measuring the span, that is, by increasing the number of objects progressively until the participant could not recall a list perfectly. A second goal was to make this experiment the hardest possible by preventing chunking processes. Our hypothesis was that participants could not reach a span greater than the magical number 3/4 (Cowan, 2001) in this condition.

Participants. Ninety five students enrolled at the University of Franche-Comté or the University of Nice, $M = 21.4$ years old ($sd = 3.7$), volunteered to participate in this quick experiment.

Procedure. The experimental setting was similar to Experiment 1, except that the number of objects in the list progressively increased until supraspan conditions were reached. There were two different lists of objects by list length. The experiment automatically stopped after two consecutive errors. The most heterogeneous categories (i.e., noncompressible categories) were chosen for each list length and a dissimilar order was applied to the list so that the overall distance between the objects would be maximal for a given list. For instance we used the categories for which FC = 6 when two objects were used; FC = 8 for three objects; FC = 10 for four objects, etc.

The participants were again asked to recall the objects in order. The experiment lasted on average around 5 minutes. The procedure did not include any warm-up.

Scoring. A .5 value was cumulated for each perfectly correct serial report of a list. For instance, a participant who correctly recalled two lists of one object, two lists of two objects, and only one list of three objects was granted a span of 2.5.

Results

The experiment was found very difficult for participants, which is confirmed by the results. Figure 10 shows the histogram of the span. A mean span of 2.5 (std = 0.6) was found across participants, a value slightly inferior to the 3/4 capacity limit. This measure seems however quite stable since we had obtained the exact same estimate from our first sample of 60 students from the Université de Franche-Comté (a sample size that we had already increased at that time in order to confirm the low span).

Discussion

Short-term and long-term memory. The chunking memory span task presented in this study provided a way to observe the birth of a chunk in working memory. This task required the participants to memorize diverse sequences of objects with possible associations between them, but the associations were induced by informativeness instead of being spontaneously created by participants (K. Z. Li, Blair, & Chow, 2010; Logan, 2004). Also, chunking effects were induced within-sequences, without relying on rendering chunking more familiar via repeated exposure (Majerus, Perez, & Oberauer, in press), a procedure that invited participants to chunk information 'on the fly' (a process that could be assimilated to grouping, as discussed later). When Feldman's (2000) rules of classification were used to measure the compressibility of each sequence of stimuli, the compressibility of the sequences and their presentation order both showed significant influences on recall performance and response times. Recall performance greatly varied around a mean span of about 4 objects or less (at 50% corect), depending on whether chunking could lower memory demand by data compression. These variations in performance mean that chunking processes can operate during short-term memorization. The first conclusion is therefore that chunks must not be uniquely considered as groupings of information in long-term memory. We believe that we observed a chunking process that seems to result from the temporary creation of new logical abstraction, or in other terms, that this chunking process does not necessarily result from the recruitment

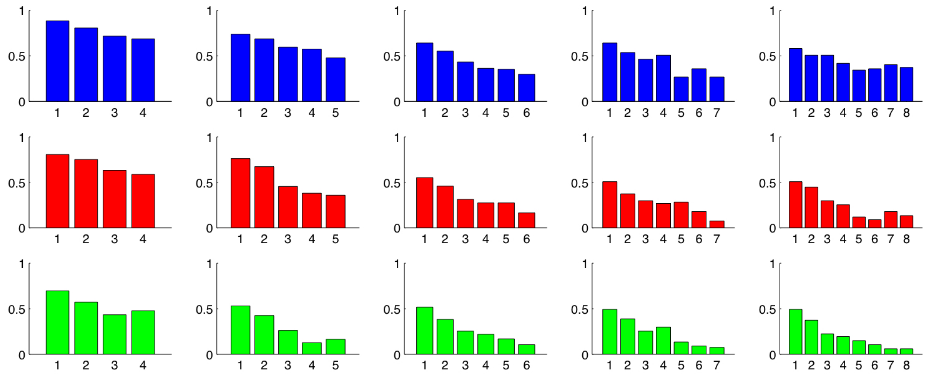

Figure 7. Serial-position curves (mean proportion correct) as a function of list length and presentation order. The blue, red and green colors stand for Rule-based, Similarity-based, and Dissimilarity-based respectively, to follow Fig. 5.

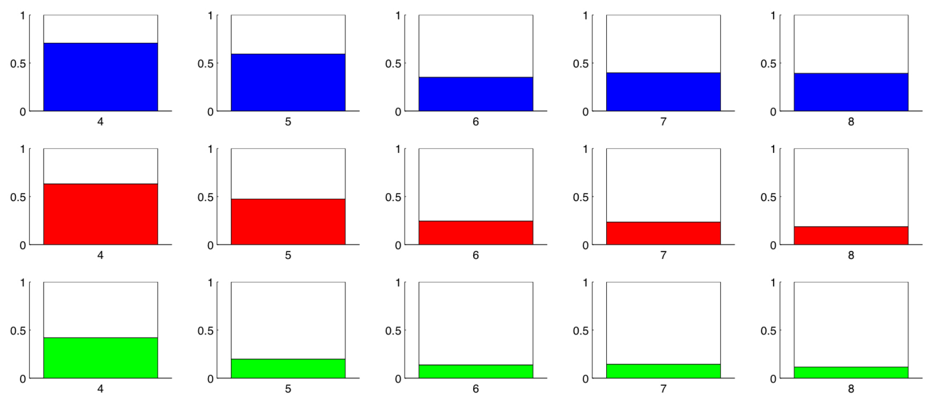

Figure 8. Mean proportion of correct recalled two-object pairs using strict serial-position scoring (pairs in positions 1-2, 3-4, 5-6, 7-8) as a function of list length and presentation order. The blue, red and green colors stands for Rule-based, Similarity-based, and Dissimilarity-based respectively, to follow Fig. 5.

of long-term memory structures. It is possible that our participants spontaneously compressed information without a strong mediation of long-term memory chunks, since none of our participants probably ever encountered the particular sequences of objects of our experiment. Memory chunking probably emerged in short-term memory simply thanks to the conscious identification of structures in the input stream. This directly relates to the idea that chunking processes are mediated by consciousness (Bor & Seth, 2012), especially when subjects learn rule-based category structures via explicit processes (Ashby & Ell, 2001). Therefore, the chunking memory that our tasks refer to does not exactly correspond to the notion of conceptual short term memory developped by Potter, which operates by allowing meaningful patterns in long term memory to be identified in short term memory unconsciously and without the need to rehearse the material (Potter, 1993, 2012).

Capacity. Our results corroborate the hypothesis of a limited capacity of short-term memory of about 3.5 items at 50% correct recall, a capacity limit that is close to the one evaluated by Cowan, about 4 ± 1 items. The capacity is also close to an estimation made by Brady, Konkle and Alvarez (2009) who suggest that short-term memory capacity is about 10 bits of information for this type of material. Since each of the stimuli making up our material counted three informational units – shape, color and size -, retaining 3.5 items corresponds to a capacity to code and recall 10.5 bits of information. However, this computation is made by considering that the recall of objects was independent, which was not the case since the effects of chunking were present. Consequently, the 3.5 limit could also confirm the observations of Luck and Vogel (1997) that individuals retain objects

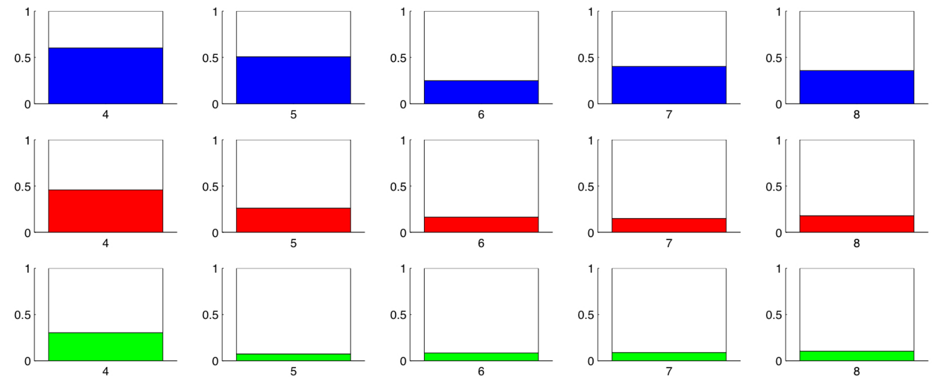

Figure 9. Mean proportion of perfectly correct recall of the first quadruplet (items in positions 1-4) as a function of list length and presentation order. Note. The slight nonmonotonicity caused by the condition in which there were 6 to-be-recalled objects seems to have been caused by the categories for which FC = 6, in which the first quadruplet was difficult to encode in comparison to the case for which FC = 5. Greater monotonicity (as a function of list length) was observed when this condition was removed from the analysis. The blue, red and green colors stands for Rule-based, Similarity-based, and Dissimilarity-based respectively, to follow Fig. 5.

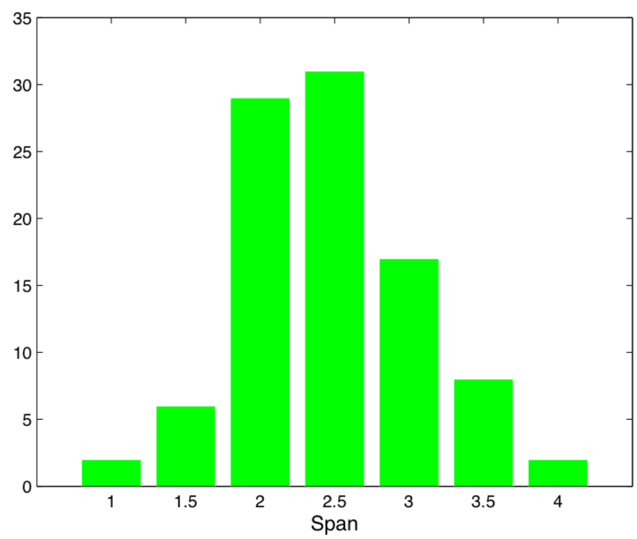

Figure 10. Mean span for Experiment 2. Note. Objects lists were designed to prevent chunking.

and not features (otherwise, again, the span would present a much higher value, that is $3.5 \times 3 = 10.5$ features), although we acknowledge that this claim tends to ignore the complex literature that directly examines this question (Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Brady et al., 2011; Fougnie & Alvarez, 2011; Wheeler & Treisman, 2002). More importantly, the 3.5 limit confirms that our study on chunking concerns STM processes.

The results of the present study tend to agree with an intermediary point of view that the informational load determines the number of objects which can be retained (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Brady et al., 2009). More objects can be unpacked when more information can be compressed within chunks. However, to be more precise, the difference between our study and previous ones (e.g., Alvarez & Cavanagh, 2004) is that the informational load does not concern items but rather relational information about individual items. In this respect, our primary goal involved characterizing the ability of observers to remember more items that allow encoding joint-information (Jiang, Olson, & Chun, 2000). Our results show that items of the memory sequence are not represented independently. Effectively, for any sequence of N objects, the sum of the visual informational load was constant in our study when the items were considered independently. However, the information load varied according to how the set of items could be encoded. This is why we believe our study directly taps into chunking processes, that is, the ability to group items so as to reduce the load of information. Moreover, chunks (such as FBI) are often seen as allowing the retrieval of their content. Our view is similar here since a group of objects can be unpacked from a chunk (e.g., squares = "large black , small black square, etc.". Chunking in our study was neither the result of perceptual grouping (such as spatial proximity) nor the result of knowledge stored in long-term memory (such as F+B+I = FBI), but the effect of a conceptual organization that could be built in a few seconds.

Representation formats. We believe that our results show that the mechanism by which compressibility by itself helps retention of one ordered list is clearly determined by the chunk that can be formed using order information, which we have represented using embedded structures such as "triangle[white:...[small:...]]". This structure means that the triangle objects have to be listed by color (beginning with the white), and that a small:large order is embedded into the color order. Note that the "triangle[white:...[small:...]]" representation is short for "triangle[white:black[small:large]]", since there exists only two features by dimension, '...' is the default opposite feature. Listing all the white triangles using the small:large order, for instance, would rather require a representation such as "triangle[white[small:...]]" or simply "white triangle[small:...]]".

The "triangle[white:...[small:...]]" compressed representation can be decompressed into a "triangle/white/small, triangle/white/large, triangle/black/small, triangle/black/large" ordered list, which comprises the original 12 features. When no such compression is possible (e.g., (large white or small black) square or (large black or small white) triangle; when FC = 10), the representation must contain almost as many features as in the original list of objects. On the contrary, when the category for which FC = 1 is ordered in a way that does not allow regularity, the memorization of all objects –unordered– can be facilitated by its global compressibility ("triangles"), unlike the memorization of the ordered objects (e.g., large white triangle, small black triangle, small white triangle, large black triangle), which still need to be expressed using a longer representation such as triangle[large white, small[black:...], large black]. Therefore, a compressible category that is not ordered regularly can be as difficult to memorize as a noncompressible one. This is noticeable in Fig. 5 in which the mean data point for FC = 10 (black curve) is very close to the mean data point for the dissimilarity green curve. Again, when no compressibility is allowed in the first place, order does not matter much anymore. Note that this idea implies that the items in a list formed from a compressible category require shorter verbal descriptions, which can be used to encode the list.

Short-term memory vs working memory. More importantly, we think that the chunking memory span tasks developed in this study and in a previous one (Mathy & Feldman, 2012) offer a valuable option to the WM/STM distinction (Colom, Rebollo, Abad, & Shih, 2006; Engle, 2002; Oberauer, 2009). STM is primarily concerned with storage. WM is most often characterized by a system dedicated to both short-term storage and the processing of information. This view has been translated in the complex span paradigm, in which the storage of memory items alternates with processing the not-to-be-remembered material. On the contrary, our chunking memory span tasks require the participants to manipulate the to-be-remembered material. Finally, the strong relationship found between STM and WM and their common variance with more general cognitive abilities (Colom et al., 2006) could be interpreted otherwise if WM was compared to chunking memory. Generally, the concurrent task directs the focus of attention away from the memory items. Here, the focus of attention promotes the grouping, chunking, and direct memorization of items. The to-be-tested idea is that the processing component plays a more major role when dedicated to memory items, which can lead to a better prediction of complex cognitive abilities such as reasoning, intelligence, or academic performance. Effectively, a weak storage component can be supported by a powerful processing component. Overall, the storage-processing combination might be the most correct composite variable for predicting cognitive abilities. Further analysis shall involve, for instance, studying the extent to which STM, WM and chunking memory spans are related to each other. We have recently completed a study in our lab, which tends to show that performance in chunking memory span tasks for alphanumerical material correlate better with WM spans than simple letter spans do, especially for memory updating tasks that also require to process the to be-remembered items (Mathy & Varré, 2013).

Compressibility of information. In this study, the measure of compressibility of information is based on the work of Feldman (2000) who rigorously quantified a metric that predicts psychological difficulty in categorization processes. The metric of Feldman makes it possible to evaluate the cost of encoding, when it is understood that complexity is not sensitive to the relative choice of a description language (M. Li & Vitányi, 1997). In the present study, recall performance was significantly enhanced for compressible sequences compared to less compressible sequences. If new categories were effectively dynamically made on the fly, it would mean that categorization complexity applies at the very beginning of the learning process. This result can be explained by a reduction in the quantity of information encoded that results from the individuals' capacity for abstraction in a short period of time, with no necessity for long-term memory processes to operate (French, Addyman, & Mareschal, 2011). Our results tend to reinforce the validity of Feldman's metric. Effectively, in a general manner, whatever the length of the sequence, the higher the complexity index, the lesser the probability of chunking (i.e., performance decreases in a significant way). The effect is more pronounced if performance is measured at a constant length as a function of logical complexity: performance diminishes significantly when logical complexity increases. Concerning this point, we agree with Brady, Konkle and Alvarez (2009) who suggest that chunking can be used as an approximation of an algorithm of psychological compression. More importantly for the working memory domain, our study tends to show that the short period of time allowed in our tasks (one- or two- second presentations of the memory items) is sufficient to allow a process of lossless abstraction to occur (when recall is perfect). This result adds to a previous observation that memory for individual items can also be biased by a lossy form of abstraction (Brady & Alvarez, 2011).

An important point is that compressibility can help item memory without having all the items to be held in memory already. Effectively, the lossless compression process allows the objects to be recovered from the compressed information. For instance, a chunk such as "square:[large:[black:]]" can be regarded as a shorter description (i.e., a definition) of the category defined in extension by "large black square, large white square, small black square, small white square, large black circle, large white circle, small black circle, small white circle". Absolutely no information is lost if the shorter representation serves as a generator for ordering hierarchically all the square shapes first, beginning with the large ones and, among the large ones, always beginning with the black ones. This resembles an algorithmic structure in which three loops are embedded: for shape = square : circle, for size = large : small, for color = black : white, concatenate (shape,size,color). The shorter representation can therefore be thought as a pseudo algorithm. On the contrary, there is no way to compress a category such as "large black square, small white square, large white circle, small black circle", which requires a more extensive description in STM.

A partial correlation analysis was carried out to show that at a constant length, compressibility and correct proportion are negatively linked. This was intended to go beyond the potential confound between compressibility and category length due to item order: long sequences can be highly compressible. For instance, the two categories on the second line of Figure 3 have equivalent Feldman Complexity, but recovering the order of all the "small or square" objects in comparison to the "large triangles" is necessarily more difficult. Some readers might argue that another confound relates to the possibility of making less item errors with longer sequences (i.e., with length 8, there was no possibility of making an item error). As a result, difficulty with longer sequences was simply underestimated in our study. Our procedure could have added a number of lures proportional to the length in the probe image in order to have probe set sizes always superior to the memory set sizes, for instance adding a set of 8 different objects in the test window. However, in our opinion, the extra objects (for instance, the extra objects could be distinguished by an extra dimensional value, such as "hatched" instead of plain) would be easily excluded from the list of the potential items to be recalled. We admit that not all of our data have the same value for testing our ideas, but again, our goal was to study how chunking can occur in WM, focusing on the memorability of different sequences of the same sequence length, rather than focusing on the decrease in performance with sequence length, a topic that has already been studied extensively.

Presentation order. The manipulation of presentation orders in the sequences aimed to explore the nature of the process of abstraction used by the participants. The regularity of one order improves the chance of noticing the to-be-compressed regularities. Our hypothesis was that the reduction of information is done by abstracting logical rules, rather than by forming similarity associations. Following Mathy and Feldman (2009, submitted), we constructed three order conditions in making the hypothesis that an order by rule would favor the compression of material and its subsequent recall in the correct order. Conforming to our expectations, the effect of the order conditions on recall performance showed better performance when the items were presented by rule, in distinguishable clusters. Thus, in the Rule condition, recall was higher than in the Similarity condition, and even higher than in the Dissimilarity condition.

We believe that the models of short-term memory based on the distinctiveness of stimuli do not explain these presentation order effects. These models consider that similar items interfere between themselves and limit the possibility to retain them in a sufficiently distinct way in order to recall them correctly (Hasher & Zacks, 1988; Lewandowsky, Duncan, & Brown, 2004; Nairne, 1990, 2002; Oberauer & Kliegl, 2006). Consequently, these models could predict better recall in the Dissimilarity condition (this can also be the case in studies on visual STM: Avons & Mason, 1999, p. 223) given that the dissimilarity is increased between contiguous pairs of objects within those lists (although this prediction would not hold when taking into account the overall within-list similarity between items, which would be the same across presentation orders in our tasks). However, the process of chunking can explain the better recall with consecutive similar stimuli: the closest shapes were grouped, treated in chunks, and therefore they allowed an increase in capacity for recall. The increase in capacity is then explicable by a compression of information which can result from grouping processes, easier and faster to do when the presentation of the objects is organized by similarity or clusters. This result supports a more general finding that similarity can lead to better retention in visual short-term memory (Johnson, Spencer, Luck, & Schöner, 2009; Lin & Luck, 2009) or verbal working memory (Gupta, Lipinski, & Aktunc, 2005; Nairne & Kelley, 1999).

Nevertheless, similarity as the exclusive criterion of order between successive objects led to worse performance than in the Rule condition. In the Rule condition, regularities were in part associated with similarities in the presentation order. However, the similarity between two successively presented objects was low between two clusters. Moreover, the clusters of a same sequence grouped pairs of objects in the same order. In conformity with the results of Mathy and Feldman (2009), we believe that the rule-based order led to higher performance by favoring the formation of rule abstractions. This result also goes in the direction of work on free recall showing that similar items tend to be recalled together even if they are not presented together (Bousfield & Cohen, 1953; Hintzman, Block, & Inskeep, 1972; Kruschke, 1996; Lewandowsky, Brown, & Thomas, 2009; Romney, Brewer, & Batchelder, 1993). The reorganization of recall is facilitated here, because the items are grouped beforehand during the presentation.

Encoding time. The participants in the 2s condition showed significantly better recall performance than those in the 1s condition, although the effect size was low. In the 2s condition, the amount of time between the beginning of the presentation (displaying the first item) and the beginning of the recall phase was doubled. Nevertheless, recall was the more effective in the 2s condition. This result does not go in the direction of a single mnemonic deterioration due to time (Baddeley, 1986; Barouillet, Bernardin, & Camos, 2004; Barouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Burgess & Hitch, 1999; Henson, 1998; Jonides et al., 2008; Nairne, 2002; Page & Norris, 1998) since more time allowed better recall. This result is explained by the difference in processing time allotted to the chunking process.

Chunking vs Grouping

Our use of the term chunking differs from the definition adopted by most researchers. The chunking memory span terminology that we have favored in our previous research (Mathy & Feldman, 2012; Mathy & Varré, 2013) emphasizes that a chunking process can be induced in a short term time by allowing associations to be made between items that present regularities. On one hand, similar manipulations (e.g., induced associations between pairs of words, temporal or spatial grouping) that have been made in the serial recall literature refer to a grouping process (Anderson & Matessa, 1997; Frankish, 1985; Hitch et al., 1996; Henson et al., 1996; Ng & Maybery, 2002; Maybery et al., 2002; Ryan, 1969). Anderson & Matessa (1997) for instance state that "knowledge units are called 'chunks' by Anderson (1993)", but that they "have repressed this terminology to avoid confusion with the term chunk as it is used in a different sense in the serial memory literature" (footnote p. 730). On the other hand, it is also true that the 'grouping' notion is ambiguous as it is also used (Feldman, 1999)[4] in the perception domain to refer to a low-level process that allows a set of visual objects to be organized into patterns, and which is assumed to be automatic and preattentive, whereas grouping in the serial recall literature is most often conscious and deliberate. Also, by using our chunking memory span terminology, we wish to introduce a differentiation between a grouping process by which groups of objects are segregated (spatially or temporarily) and a chunking process by which associations are made within groups. Our experiments effectively manipulate within-group associations instead of between-group separations. The only difference with previous studies (working memory: (Chen & Cowan, 2005; Cowan et al., 2004) ; grammar learning: (French et al., 2011; Servan-Schreiber & Anderson, 1990)) is that associations are expected to be made 'on the fly', a short-term process that does not necessarily necessitate repetition. Future studies may show that the birth of chunks that we observe in our experiments (when learning is not involved) could effectively lead to stronger associations and better chunking if lists of objects were presented more than once (i.e., when learning would be involved). One clear example of such continuity is shown by Cowan et al. (2004) in their first experiment, in which the recall of a pair of words learned only once was already better recalled on average than the singletons (and as expected, recall was a monotonic function of repetition). Therefore, although arguable, we chose to keep the chunking terminology because 1) it can make sense that there is not always a fundamental difference between short-term memory and long-term memory (Surprenant & Neath, 2009), and 2) we wished to remain consistent with our previous publications'

Conclusion

Our objective was to measure the capacity of short-term memory while allowing chunking processes to occur. We also aimed at dealing with the well-known issue of defining and measuring chunks by taking advantage of earlier research on learnability and compressibility. The experiment presented in this paper allows us to draw the conclusion that chunking is not only a matter of using representations that are familiar from long-term experience, but that chunks can be

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

formed in working memory in a few seconds by capitalizing on the compressibility of information. Our research shows that the notion of compression is central to define chunking, and also that independently from the redundancy that makes the sequences of objects more compressible, the participants benefit from other encoding possibilities depending on the order in which objects are presented. As a consequence, the sequences offering the most logical regularity are the easiest to retain, especially when the abstraction of the regularities is favored by the ordering of objects within sequences.

References

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology*, *3*(301).

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.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341-380.

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204-210.

Avons, S., & Mason, A. (1999). Effects of visual similarity on serial report and item recognition. *Quarterly Journal of Experimental Psychology: Section A*, *52*, 217-240.

Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *14*, 575-589.

Barouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83-100.

Barouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570-585.

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.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851-854.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*, 1622-1631.

Bor, D., & Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Frontiers in Psychology*, *3*(63).

Bousfield, W. A., & Cohen, B. H. (1953). The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *Journal of Psychology: Interdisciplinary and Applied*, *36*, 67-81.

Brady, T. F., & Alvarez, G. A. (2011). 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. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*, 487-502.

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

Broadbent, D. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory.* New York: Wiley.

Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.

Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551-581.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, *33*, 1-16.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1235-1249.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: a reanalysis of key studies. *Memory & Cognition*, *34*, 158-171.

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

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51-57.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to miller (1956). *Psychological Science*, *15*, 634-640.

Dirlam, D. K. (1972). Most efficient chunk sizes. *Cognitive Psychology*, *3*, 255-359.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, *7*, 149-158.

Elio, R., & Anderson, J. R. (1981). Effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 397-417.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*, 20–30.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23.

Ericsson, K. A., Chase, W., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, *12*(4), 360-372.

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 128 - 145.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological review*, *119*, 223-271.

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

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *Journal of Vision*, *11*, 1-12.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*, 200-209.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*, 614-636.

Gagné, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, *40*, 61-73.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four...or is it two? *Memory*, *12*, 732-747.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236-243.

Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain and cognition*, *79*, 221-244.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: the roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*, 1001-1016.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, *16*, 70-76.

Halford, G. S., Wilson, W. H., & Phillips, W. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803-831.

Hasher, L., & Zacks, R. T. (1988). The psychology of learning and motivation: Advances in research and theory, vol 22. In G. H. Bower (Ed.), (Academic Press ed., Vol. 22, p. 193-225). San Diego, CA: Academic Press.

Henson, R. N. A. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology*, *36*, 73-137.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*, 80-115.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning & Verbal Behavior*, *11*, 741-749.

Hitch, G., Burgess, N., Towse, J., & Culpin, V. (1996). Temporal grouping effects and working memory: The role of the phonological loop. *Quarterly Journal of Experimental Psychology*, *49A*, 116-139.

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *The Quarterly Journal of Experimental Psychology*, *25*, 104 - 111.

Hu, Y., & Ericsson, K. A. (2012). Memorization and recall of very long lists accounted for within the long-term working memory framework. *Cognitive Psychology*, *64*, 235-266.

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

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, *20*, 568-577.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193-224.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*, 225-247.

Lewandowsky, S., Brown, G., & Thomas, J. (2009). Traveling economically through memory space: Characterizing output order in memory for serial order. *Memory & Cognition*, *37*, 181-193.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review*, *11*, 771-790.

Li, K. Z., Blair, M., & Chow, V. S. (2010). Sequential performance in young and older adults: evidence of chunking and inhibition. *Aging Neuropsycholy, and Cognition*, *17*, 270-295.

Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York, NY: Springler Verlag.

Lin, P. H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Vis cogn*, *17*, 356-372.

Logan, G. D. (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General*, *133*, 218-236.

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

Macgregor, J. N. (1988). The effects of order on learning classifications by example: heuristics for finding the optimal order. *Artificial Intelligence*, *34*, 361-370.

Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 202-212.

Majerus, S., Perez, M., & Oberauer, K. (in press). Distinct long-term memory mechanisms contribute to verbal short-term memory. *Journal of Memory and Language*.

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*, 1050-1057.

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

Mathy, F., & Feldman, J. (submitted). Presentation order effects on category learning and category generalization.

Mathy, F., & Varré, J. S. (2013). Retention-error patterns in complex alphanumeric serial-recall tasks. *Memory*.

Maybery, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*, *47*(3), 360-385.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*, 307-325.

Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*, 250-254.

Medin, D. L., & Schaffer, M. (1978). A context theory of classification learning. *Psychological Review*, *85*, 207-238.

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.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.

Nairne, J. S. (2002). Remembering over the short-term: the case against the standard model. *Annual Review of Psychology*, *53*, 53-81.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*, 45-53.

Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capacity. *Memory & Cognition*, *35*, 724-737.

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: is position coded temporally? *Quarterly Journal of Experimental Psychology*, *55*, 391-424.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*, 346-357.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601-626.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, *64*, 1055-1067.

Page, M. P., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review*, *105*, 761-781.

Perlman, A., Pothos, E. M., Edwards, D. J., & Tzelgov, J. (2010). Task-relevant chunking in sequence learning. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 649-661.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General*, *119*, 264-275.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, *21*, 156-161.

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology*, *3*(113).

Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 1-19.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*, 28-34.

Ryan, J. (1969). Grouping and short-term memory: different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*, 137-147.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592-608.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 13, whole No. 517.

Simon, H. A. (1974). How big is a chunk? *Science*, *183*, 482-488.

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual—spatial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology: Section A*, *58*, 909-930.

Surprenant, A. M., & Neath, I. (2009). Interactions between short-term and long-term memory in the verbal domain. In A. Thorn & M. P. (Eds.) (Eds.), (p. 16-43). Hove, UK: Psychology Press.

Terrace, H. S. (1987). Chunking by a pigeon in a serial learning task. *Nature*, *325*, 149-151.

Terrace, H. S. (2001). Chunking and serially organized behavior in pigeons, monkeys and humans. In R. G. Cook (Ed.), *Avian visual cognition.* [On-line]. Medford, MA: Comparative Cognition Press, available at: www.pigeon.psy.tufts.edu/avc/terrace/.

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Experimental Psychology*, *16*, 83-95.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48-64.

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *The Quarterly Journal of Experimental Psychology*, *19*, 97-102.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 458-468.

Zhang, G., & Simon, H. A. (1985). Stm capacity for chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, *13*, 193-201.

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233-236.