



HAL
open science

Compression in Working Memory and Its Relationship With Fluid Intelligence

Mustapha Chekaf, Nicolas Gauvrit, Alessandro Guida, Fabien Mathy

► **To cite this version:**

Mustapha Chekaf, Nicolas Gauvrit, Alessandro Guida, Fabien Mathy. Compression in Working Memory and Its Relationship With Fluid Intelligence. *Cognitive Science*, 2018, 42, pp.904 - 922. 10.1111/cogs.12601 . hal-01873321

HAL Id: hal-01873321

<https://univ-rennes2.hal.science/hal-01873321v1>

Submitted on 13 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Compression in Working Memory and Its Relationship With Fluid Intelligence

Mustapha Chekaf,^a Nicolas Gauvrit,^b Alessandro Guida,^c Fabien Mathy^a

^a*Bases Corpus Langage UMR 7320 CNRS, Université Côte d'Azur*

^b*Human and Artificial Cognition Lab, EPHE*

^c*LP3C, University of Rennes*

Received 25 August 2016; received in revised form 23 October 2017; accepted 21 January 2018

Abstract

Working memory has been shown to be strongly related to fluid intelligence; however, our goal is to shed further light on the process of information compression in working memory as a determining factor of fluid intelligence. Our main hypothesis was that compression in working memory is an excellent indicator for studying the relationship between working-memory capacity and fluid intelligence because both depend on the optimization of storage capacity. Compressibility of memoranda was estimated using an algorithmic complexity metric. The results showed that compressibility can be used to predict working-memory performance and that fluid intelligence is well predicted by the ability to compress information. We conclude that the ability to compress information in working memory is the reason why both manipulation and retention of information are linked to intelligence. This result offers a new concept of intelligence based on the idea that compression and intelligence are equivalent problems.

Keywords: Intelligence; Short-term memory; Working memory; Human information storage; Chunking

1. Introduction

Although there is no doubt that working memory (WM) is closely related to fluid intelligence (Kane, Hambrick, & Conway, 2005), there is no definite answer as to why manipulation and retention of information are responsible for variations in individual fluid intelligence (henceforth referred to as Gf). The main account describing this relationship reported so far is based on memory capacity (Oberauer, Süß, Wilhelm, & Sander, 2007).

Correspondence should be sent to Fabien Mathy, Département de Psychologie, Université Côte d'Azur, Laboratoire BCL: Bases, Corpus, langage – UMR 7320, Campus SJA3, 24 avenue des Diabls Bleus, 06357 Nice CEDEX 4, France. E-mail: fabien.mathy@unice.fr

For instance, the role of capacity increases with items' difficulty of Raven's Progressive Matrices (Little, Lewandowsky, & Craig, 2014). Although WM is known to play an indisputable role in intelligent behaviors, our goal is to go further to shed light on the process of information compression in WM as a determining factor of intelligence. Compression of information, that is, the capacity to recode information into a compact representation, may be a key mechanism in accounting for the dual impact of the manipulation and retention of information on intelligence, particularly when individuals deal with new problems in the typical tests used to measure intelligence (such as Raven's matrices).

Compression can potentially account for intelligence because many complex mental activities still fit well into rather low WM storage capacity—for similar ideas, see Baum (2004) or Hutter (2005) in artificial intelligence, and Ferrer-i-Cancho et al. (2013) in behavioral science. Our goal here is to show that the ability to compress information (see Brady, Konkle, & Alvarez, 2009) is a good predictor of WM performance and is, thereby, a good candidate for predicting intelligence: More compression simply means more available resources for processing other information, which can potentially have an impact on reasoning. For instance, Unsworth and Engle (2007) showed a better prediction of fluid intelligence after increasing the length of the memoranda. They observed that when list length reached five items, simple spans became as good as complex spans in terms of predicting fluid intelligence. One possible explanation is that long-length lists need to be reorganized by individuals to be stored.

The memory span revolves around four items when the task is complex (i.e., rapid or dual; see Cowan, 2001), whereas the span revolves around seven items when the span task is simple (e.g., when the participant is only required to repeat back a sequence of simple items such as letters or digits; Miller, 1956). Therefore, there is a paradox in the usual STM/WM concepts, because it is difficult to conceive that the STM estimate is the highest if processing is not included in the concept of STM. Effectively, STM is conceptually thought to be primarily involved in storage; however, items can be manipulated and potentially compressed when the span task is simple. This may explain why the span can go as high as seven items on the digit-span task. Conversely, while WM is primarily involved in both the manipulation and storage of information (conceptually), the capacity to manipulate the memorandum is generally limited in complex span tasks and the measure of interest is usually storage (for an exception, see Unsworth, Redick, Heitz, Broadway, & Engle, 2009). This study proposes and develops the idea that the concept of information compression can help eliminate the opposition between the STM and WM concepts and may be useful in accounting for intelligence in a comprehensive way. Effectively, one observation is that compression processes play a role in variations of the span (Chekaf, Cowan, & Mathy, 2016) around the average 4 ± 1 WM capacity limit. In this study, the compression process allowed the formation of chunks that simplified participants' memorization. A second observation is that the previous 7 ± 2 estimation of short-term memory (STM) capacity may correspond to an overestimation due to compression of several items into larger chunks (Mathy & Feldman, 2012).

From this standpoint, our hypothesis is that the capability of simple spans (when using long-length lists and particularly when they can be compressed) to predict fluid

intelligence could be due to information reorganization. More specifically, we hypothesize that information reorganization and intelligence should both rely on individuals' storage capacity, which hierarchically depends on a compression process that can help optimize the available storage. Storage capacity is supposed to be fixed for a given individual; however, any form of structuration of the stimulus sequence can contribute to extending the number of stored items. This optimization process could be helpful to make sense of the regularities that can be found in the items of the Raven or in the to-be-recalled sequences of repeated colors. The rationale is that humans use storage and compression in conjunction to solve the problems of the Raven and to reorganize (for instance, by chunking) the colors of the to-be-remembered sequences of repeated colors.

To study the ability to compress information, we developed a span task based on the SIMON[®], a classic memory game from the 1980s that can be played on a device with four colored buttons (red, green, yellow, and blue). The game consists of reproducing a sequence of colors by pressing the corresponding buttons. A similar task has proved successful to measure the span in previous studies (Gendle & Ransom, 2009; Karpicke & Pisoni, 2004). Our SIMON span task was developed on the reasoning that sequences of colors contain regularities that can be quantified to approximate compression opportunity. To estimate the compressibility, our participants could be offered within thousands of different color sequences in the SIMON task, we used a compressibility metric that provides an estimate of every possible rapid grouping process that can take place in WM. To anticipate, the present results showed that this metric can indeed be used to predict WM performance and that intelligence is well predicted by the ability to compress information. We conclude that the ability to compress information in WM is the reason why both the manipulation and the retention of information are linked to intelligence.

2. Measuring complexity or compressibility

To estimate the ability to compress information, the objective is to measure how much a given string s (of colors, in our case) can potentially be shortened by a lossless recoding process.¹ This can be achieved as soon as we can estimate the length of the shortest possible description of s . This length has been estimated in various ways in the past; however, all methods provide a definition of "complexity." Effectively, the complexity of a string s can be defined as the length of the shortest possible description or, equivalently, of the shortest program that would produce the string and then halt. It is, thus, equivalent (in a reverse manner) to the compressibility we seek to estimate, with the general idea that the simplicity of an object is measured by the length of its shorter description (Chatter & Vitányi, 2003). Note that we do not intend here to describe how humans recode but only to estimate their ability to compress. Thus, our aim was not to yield an entirely psychologically plausible compression process but rather to measure theoretically a compression rate.

Past research has used various methods to estimate complexity. For instance, some used a simple intuitive index of complexity of the string based on the number of

repetitions (Brugger, Monsch, & Johnson, 1996), or based on the type-token-ratio (TTR) defined as the ratio of the number of different colors by the length of the string (Manschreck, Maher, & Ader, 1981). Others have used more sophisticated methods such as Shannon's entropy (Gauvrit, Soler-Toscano, & Guida, 2017; Pothos, 2010) or minimum description length (Fass & Feldman, 2002; Robinet, Lemaire, & Gordon, 2011). For a string s , entropy $H(s)$ is defined as follows:

$$H(s) = - \sum_i p_i \log_2(p_i),$$

where i are the different possible symbols (here, colors) and p_i the proportion of symbol i in the sequence (Shannon, 1948). As seen from this formula, the entropy of a string only depends on the relative proportions of symbols and is unaffected by the order in which they appear. For instance, the entropy is maximal whenever the different symbols appear with equal probability, irrespective of their layout. For instance, the two binary strings 0010111001 and 0101010101 have the exact same (maximal) entropy. This is not only counterintuitive but also contradictory with other complexity measures such as MDL. Although entropy is mathematically sound and statistically related to other measures of complexity, it is not the appropriate measure in our case. Besides, higher-order entropy (for instance two-order entropy based on diagrams instead of unigrams) has been put forward as a way to counter the limitations of first-ordered Shannon's entropy, and they actually partly do so (Gauvrit, Singmann, Soler-Toscano, & Zenil, 2015). However, (a) some strings have maximal second-order entropy despite being intuitively simple, such as "001100110011..." with an equal number of "01," "10," "00," and "11," and (b) higher-order entropy is in practice impossible to use with very short strings such as those used in this study.

MDL (Rissanen, 1978) consists of choosing a coding language for a given s to determine its shortest definition using that language. For instance, the string 0101010101 could be described as $(01)^5$, a shorter description than the string itself, given that the chosen language includes a power operator to indicate repetitions. MDL is, thus, a natural avenue to define human compression in a plausible way as soon as we know what language best describes how humans compress. However, when the objective is to use a universal and objective normative complexity measure, MDL is too dependent on the choice of a language. Furthermore, probably more importantly here, MDL (in any of its usual version) offers no advantage for recoding short sequences such as those studied in our experiments. MDL can be clearly useful for a sequence that is quite long, such as 123123123123123123123, for which we can use a shorter representation $a = 123$, $aaaaaaa$ (the $a = 123$ part corresponds to the language used in a lookup table, and $aaaaaaa$ corresponds to the recoded sequence, and combined, " $a = 123$, $aaaaaaa$ " is shorter than the original string). However, rewriting *blue-blue-blue-red* to take into account the *blue-blue-blue* regularity can take more space than the original sequence. As a consequence, depending on the chosen language, 'blue-blue-blue-red' might not be

recorded by an MDL procedure, and thus the intuitive regularity might remain undetected.

Because the measures of complexity mentioned above focus on particular aspects of complexity or are based on assumptions about the type of compression that should be used (i.e., the chosen language in the case of MDL), it has often been concluded that the (Kolmogorov–Chaitin) Algorithmic Complexity should be used instead (Chater & Vitányi, 2003). The algorithmic complexity of a string s is defined, very much in line with the MDL principle, as the length of the shortest program that, running on a universal Turing machine (an abstract all-purpose computer), will produce s and then halt.

Although algorithmic complexity is a universal and objective measure of complexity (Li & Vitányi, 1997), it is uncomputable: There is no algorithm taking any finite string as input and giving as output its algorithmic complexity. However, the “coding theorem method”² (Delahaye & Zenil, 2012; Soler-Toscano, Zenil, Delahaye, & Gauvrit, 2013, 2014) offers a reliable approximation of the algorithmic complexity of short strings (with length from 3 to 50 symbols). Note also that contrary to compressibility measures such as MDL, Algorithmic Complexity makes no assumption about the types of algorithm or language that is preferred. The compressed version of a string (equated to the program that produces it) is only limited by the fact that the program is an algorithm. Hence, any computable method to obtain a string is relevant. For instance, it encompasses (but is not limited to) chunking, which can be understood as the decomposition of a string into shorter substrings. The coding theorem method ensures the best estimation of algorithmic complexity, specifically for short strings because a long-standing problem in this field is that short sequences of symbols (i.e., less than a 100, as in this study) are difficult to compress.² The method has already been used in psychology, in domains other than WM or intelligence (e.g., Dieguez, Wagner-Egger, & Gauvrit, 2015; Gauvrit, Soler-Toscano, & Zenil, 2014; Kempe, Gauvrit, & Forsyth, 2015), and it has now been implemented in a user-friendly manner as an R-package (Gauvrit et al., 2015).

3. Method

Our goal was to devise a new span task enabling us to measure the human ability to compress information in WM. Concurrent measures of WM were administered for comparison with the new span task and its predictions of Gf.

3.1. Participants

One hundred and eighty-three students enrolled at the University of Franche-Comté in Besançon, France ($M_{\text{age}} = 21$; $SD = 2.8$) volunteered to participate in the experiment and received course credit in exchange for their participation.

3.2. Procedure

Depending on their availability and their need for course credit, the volunteers took one, two, or three tests: our adaptation of the game SIMON[®] (hereafter called the SIMON span task), the working memory capacity battery (WMCB), and the Raven's advanced progressive matrices (APM) (Raven, 1962). All volunteers were administered the SIMON span task; some also took the WMCB ($N = 27$), Raven's matrices ($N = 26$), or both³ ($N = 85$). In all cases, tests were administered in the following order: (a) SIMON span task, (b) Raven, (c) WMCB.

3.3. SIMON span task

Each trial began with a fixation cross in the center of the screen (1,000 ms). The to-be-memorized sequence, consisting of a series of colored squares appearing one after the other, was then displayed (see Fig. 1). Next, in the recall phase, four colored buttons were displayed, and participants could click on them to recall the whole sequence they had memorized and then validate their response. After each recall phase, feedback ("perfect" or "not exactly") was displayed according to the accuracy of the response.

Participants were administered either a spatial version or a nonspatial version of the task. In the spatial version ($N = 106$), the stimulus colored squares were displayed briefly lit up one at a time, in different locations on the screen, to show the to-be-remembered sequence, as in the original game. The spatial version was primarily run to follow the original game first; however, we also added a stringent condition (called the nonspatial

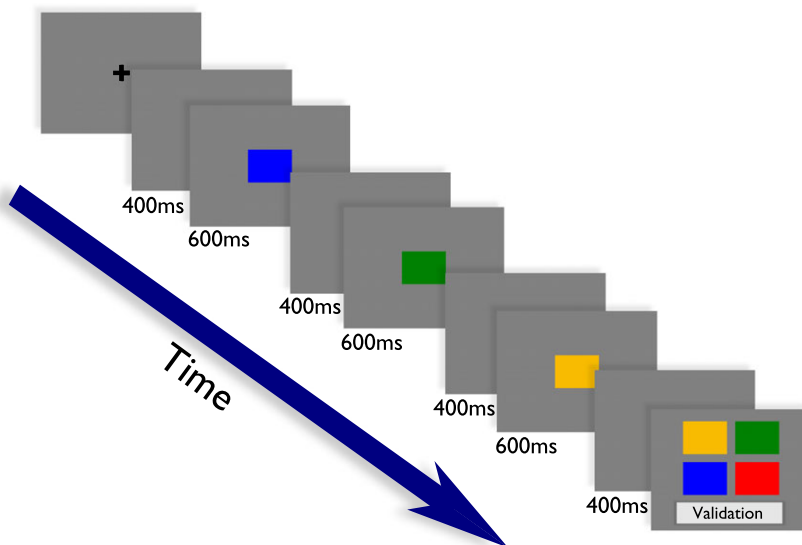


Fig. 1. Example of a sequence of three colors for the memory span task adapted from the SIMON[®] game.

version) to better control spatial encoding. In the nonspatial version ($N = 77$), the stimulus colors were displayed one after another in the center of the screen to avoid any visuospatial encoding strategy. To further discourage spatial strategies in both versions, the colors were randomly assigned to the buttons on the response screen for each trial, which resulted in the colors never being in the same locations across trials. The reason for avoiding spatial encoding was that our metric was developed for detecting regularities in one-dimensional sequences of symbols, not two-dimensional patterns. Given that our preliminary analysis indicated no significant difference between these two conditions, the two datasets were combined in the subsequent analyses reported in the Results section.

Each session consisted of a single block of 50 sequences varying in length (from one to ten) and in the number of possible colors (from one to four). New sequences were generated for each participant, with random colors and orders, so as to avoid the presentation of the items in ascending length (two items, then three, then four, etc.). We chose to generate random sequences and measure their complexity a posteriori (as described below). A total of 9,150 sequences (183 subjects \times 50 sequences) were presented (average length = 6.24), each session lasted for 25 min on average.

3.4. Working Memory Capacity Battery

Lewandowsky, Oberauer, Yang, and Ecker (2010) designed this battery for assessing WM capacity by means of four tasks: an updating task (memory updating, MU), two span tasks (operation and sentence span, OS and SS), and a spatial span task (spatial short-term memory, SSTM). It was developed using MATLAB (MathWorks Ltd.) and the Psychophysics Toolbox (Brainard, 1997). On each trial of MU, participants were required to encode an initial set of digits (between three and five across trials). Each digit was presented on the screen in a separate frame. The digits were presented one after another, for 1 s each. Participants were then required to update these digits when shown arithmetic operations such as “+3” or “-5.” These cues were displayed in the respective individual frames one at a time for 1.3 s each. Updating consisted of replacing the initial memorized digit by the result of the arithmetic operation on the digit. After a series of updating operations, final recall (of the updated set of digits) was signaled by question marks appearing in the frames.

In both OS and SS, a complex span-task paradigm was used in such a way that the participants saw an alternating sequence of to-be-remembered consonants and to-be-judged propositions: the judgments pertaining to the correctness of the equations in the OS task or the meaningfulness of the sentences in the SS task. The participants were required to memorize the complete sequence of consonants for immediate serial recall.

In SSTM, the participants were required to remember the location of dots in a 10×10 grid. The dots were presented one by one in random cells. Once the sequence was completed, the participants were cued to reproduce the pattern of dots using a mouse. They were instructed that the exact position of the dots or the order was irrelevant; the important thing was to remember the pattern made by the spatial relations between the dots. The four tasks together took 45 min on average.

3.5. Raven's Advanced Progressive Matrices

After a practice session using the 12 matrices in Set 1, the participants were tested on the matrices in Set 2 (36 matrices in all), during a timed session averaging 40 min. They were instructed to select the correct missing cell of a matrix from a set of eight choices. Correct completion of a matrix was scored one point; hence, the range of possible raw scores was 0–36.

4. Results

4.1. Effects of sequence length and number of colors per sequence

The effects of sequence length and the number of colors per sequence are indicative of memory capacity in terms of the number of items recalled, and they provide an initial idea as to whether the repetition of colors generated interference during the recall process (which would predict lower performance) or permitted recoding of the sequences (which would predict better performance). First, we conducted a repeated-measures analysis of variance (ANOVA) with sequence length as a repeated measure, and the mean proportion of perfectly recalled sequences for each participant (i.e., proportion of trials in which all items in a sequence were correctly recalled⁴) as the dependent variable. Performance varied across lengths, $F(9, 177) = 234.2$, $p < .001$, $\eta_p^2 = .9$, decreasing regularly as a function of length. Fig. 2 shows the performance curve based on the aggregated data, by participant and length, and also as a function of the number of different colors in the sequence. The three performance curves resemble an S-shaped function, as in previous studies (e.g., Crannell & Parrish, 1957; Mathy & Varré, 2013).

Next, we focused on a subset of the data to study the interaction between length and number of colors. We selected trials for which sequence length exceeded three items, and we conducted a 7 (4, 5, 6, 7, 8, 9, or 10 items) \times 3 (2, 3, or 4 different colors within a sequence) repeated-measures ANOVA. There was still a significant main effect of the length factor ($M_s = .91, .82, .64, .49, .32, .2$ and $.13$, and $SEs = .01, .01, .02, .02, .02, .01$ and $.01$ for the seven lengths, respectively), $F(6, 1092) = 560.8$, $p < .01$, $\eta_p^2 = .75$. The main effect of the number of colors per sequence was also significant ($M_s = .6, .47$ and $.42$; $SEs = .01, .01$ and $.01$), $F(2, 364) = 132.4$, $p < .01$, $\eta_p^2 = .42$. Post hoc analysis (with Bonferroni corrections for the pairwise comparisons) yielded a systematic, significant decrease between each length condition and between each number-of-colors condition. Finally, there was a significant interaction between length and number of colors, $F(12, 2184) = 5.8$, $p < .01$, $\eta_p^2 = .03$: The length effect increased with the number of colors. Although this is a coarse estimation of recoding, these results show that memorization was facilitated when sequences with a few number of colors were used, and particularly when the sequences were long.

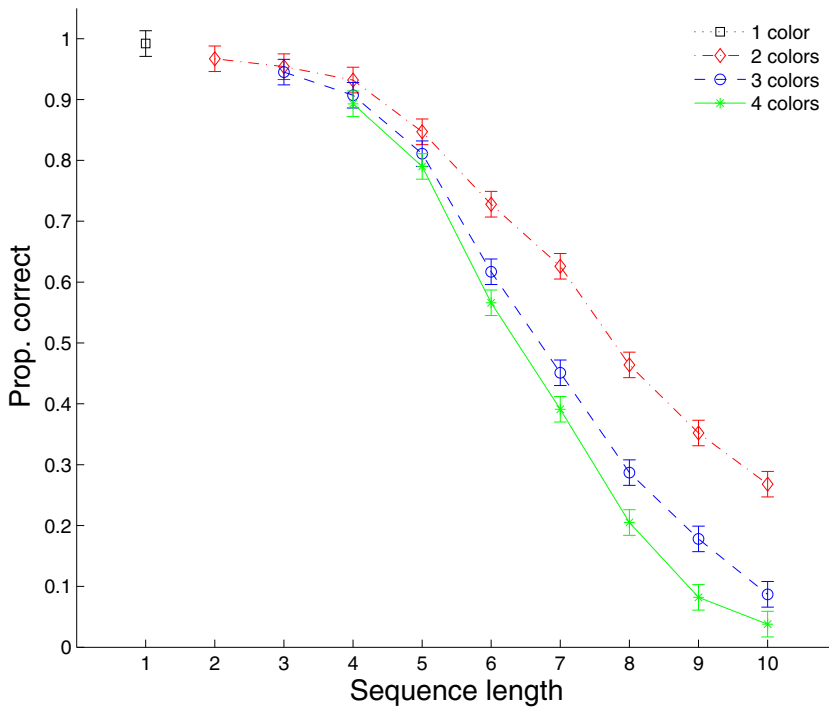


Fig. 2. Proportion of perfectly recalled color sequences, by sequence length and number of different colors within a sequence. *Note.* Error bars show $\pm 1 SE$.

4.2. Reliability of the complexity measure

To confirm the reliability of our compressibility measure (inversely related to algorithmic complexity), we split each participant's sequences into two groups of equal complexity. This split-half method simulated a situation in which the participants were taking two equivalent tests. We obtained adequate evidence of reliability between the two groups of sequences ($r = .63$, $p < .001$; Spearman-Brown coefficient = .77).

4.3. Effect of compressibility

We computed the overall link between compressibility (again, inversely related to algorithmic complexity) and the correctness of responses. The Point-biserial correlation between algorithmic complexity and correct recall of a series was $r_{pbi} = -.63$, a number that is comparable to the correlation between the length of a series and its correct recall, $r_{pbi} = -.62$. These correlations largely overtake the correlation between correct response and other index of complexity such as entropy, $r_{pbi} = -.43$, the number of colors, $r_{pbi} = -.50$, or TTR, $r_{pbi} = -.24$. This is a further justification of algorithmic complexity in our context.

A simple initial analysis targeting both the effect of the number of colors and the effect of complexity on accuracy (i.e., the sequence is perfectly recalled) showed that complexity ($\beta = -.62$) took precedence over the number of colors ($\beta = -.04$) in a multiple linear regression analysis when all 9,150 trials were included. To further investigate the combined effects of complexity and list-length on recall in greater detail, we used a logistic regression approach to predict correct performance. At each step, we chose the best predictor to add to the model. For instance, the second model is \sim Complexity because including complexity to the model produced a better fit than adding length (a BIC of 8,111.9 with complexity, but 8,480.5 with length). Then, once complexity was added to the predictors, adding length gave a better model, so we did add length. In the last stage, adding an interaction term increased the BIC, so we did not select this model. Therefore, this stepwise forward model based on a Bayesian information criterion (BIC) suggested dropping the interaction term (see Table 1). This model revealed a significant negative effect of complexity ($z(9147) = -23.84$, $p < .001$, standardized coefficient = -5.70 ; non-standardized = $-.69$), as shown in Fig. 3, and a significant positive effect of length ($z(9147) = 16.27$, $p < .001$, standardized coefficient = 3.74 ; non-standardized = 1.46). Although length had a detrimental effect on recall, this effect was largely counteracted by the detrimental effect of complexity, indicating that long, simple strings were easier to recall than shorter, more complex ones. In other words, the complexity effect was stronger than the length effect (Table 1).

Fig. 4 shows how performance decreased as a function of complexity in each length condition. The decreasing linear trend was significant for lengths 4, 6, 7, 8, 9, and 10 ($r = -.34$, $p < .001$; $r = -.39$, $p < .001$; $r = -.41$, $p < .001$; $r = -.35$, $p < .001$; $r = -.45$, $p < .001$; and $r = -.43$, $p < .001$, respectively), which shows that the memory process could indeed be optimized for the less complex sequences.

Table 1
Stepwise forward selection of logistic models based on criterion BIC

Base Model	Variable Included	Deviance	BIC
Intercept only	Complexity	8,093.7	8,111.9
	Length	8,462.3	8,480.5
	None (intercept only)	12,477.5	12,486.6
\sim Complexity	Length	7,805.2	7,832.6
	None (complexity only)	8,093.7	8,111.9
\sim Complexity & Length	None (complexity and length only)	7,805.2	7,832.6
	Interaction	7,800.0	7,836.5

Notes. The dependent variable is the response (correct/incorrect) and the factors are Complexity and Length. At each stage, for a given model (where “ $\sim v$ ” indicates that the Base model uses ‘v’ as predictor), the variables are listed in increasing BIC order after other variables were included—for instance, line 4 indicates that using both Complexity and Length yielded a BIC of 7832.6. The final model includes complexity and length, but no interaction between complexity and length. For instance, in the last model computed as a function of complexity and length, adding the interaction term increased the BIC; hence, the simplest model Complexity and Length was chosen. Both length and complexity were considered fixed effects.

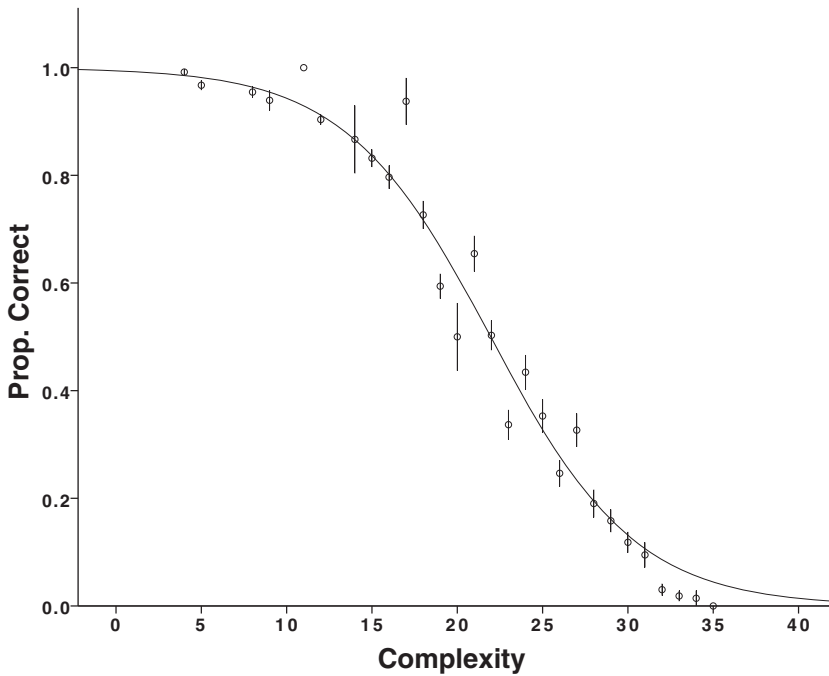
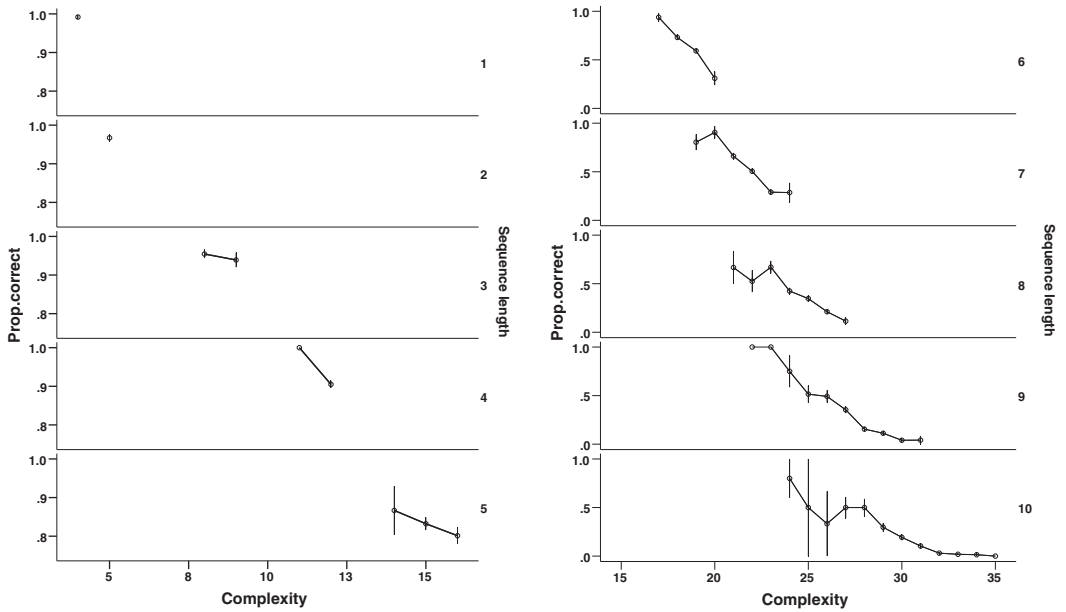


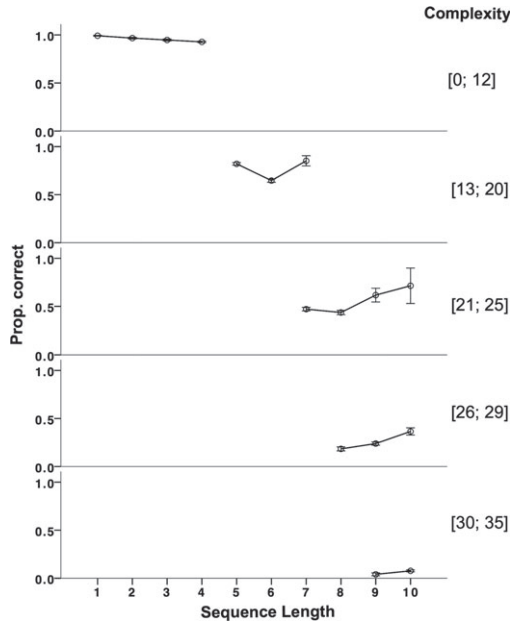
Fig. 3. Proportion of perfectly recalled sequences of colors as a function of complexity. *Note.* Error bars show $\pm 1 SE$.

One interesting result of our complexity measure was revealed in the comparison between complexity of the stimulus and response sequences. When the participants failed to recall the right sequence, they tended to produce a simple string (in terms of algorithmic complexity). The mean complexity of the stimulus sequence was 25.3; however, the mean complexity of the responses was only 23.3 across all trials ($t(9149) = 26.24$, $p < .001$, Cohen's paired $d = .27$).

Following the regression on the data shown in Fig. 3, we computed a logistic regression for each subject to find the critical decrease in performance that occurred half-way down the logistic curve (i.e., the inflection point). The reason for this was that the participants were not administered the same sequences. This inflection point is thus based on complexity and simply indicates that the participants failed on sequences more than 50% of the time when the complexity level was above the inflection point. To estimate the relationship between each participant's inflection point and his/her IQ, the data were input into confirmatory factor analysis (CFA) using IBM SPSS AMOS 21. A latent variable representing a construct in which *storage* and *processing* were separated during the task, and another latent variable, representing a construct in which the two processes functioned together, were sufficient to describe performance. The fit of the model shown in Fig. 5a was excellent ($\chi^2(7) = 2.85$, $p = .90$; comparative fit index CFI = 1.0; root mean squared of approximation RMSEA = 0.0; root-mean square residual RMR = .002; Akaike criterion (AIC) and BIC criteria were both lower than in a saturated model with all



(a)



(b)

Fig. 4. (a) Proportion of correct recall as a function of complexity, by sequence length. (b) Proportion of correct recall as a function of length, by complexity. *Note.* Error bars are ± 1 SE. There is only one type of complexity for length 2 because this condition was constrained to show two different colors.

variables correlated with each other than in an independence model with no variables correlated), with one caveat: the data failed to fit the recommended conditions for computing a CFA (see Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Wang, Ren, Li, & Schweizer, 2015) because one factor was defined by only two indicators, and also because the correlation with the two factors revealed collinearity.

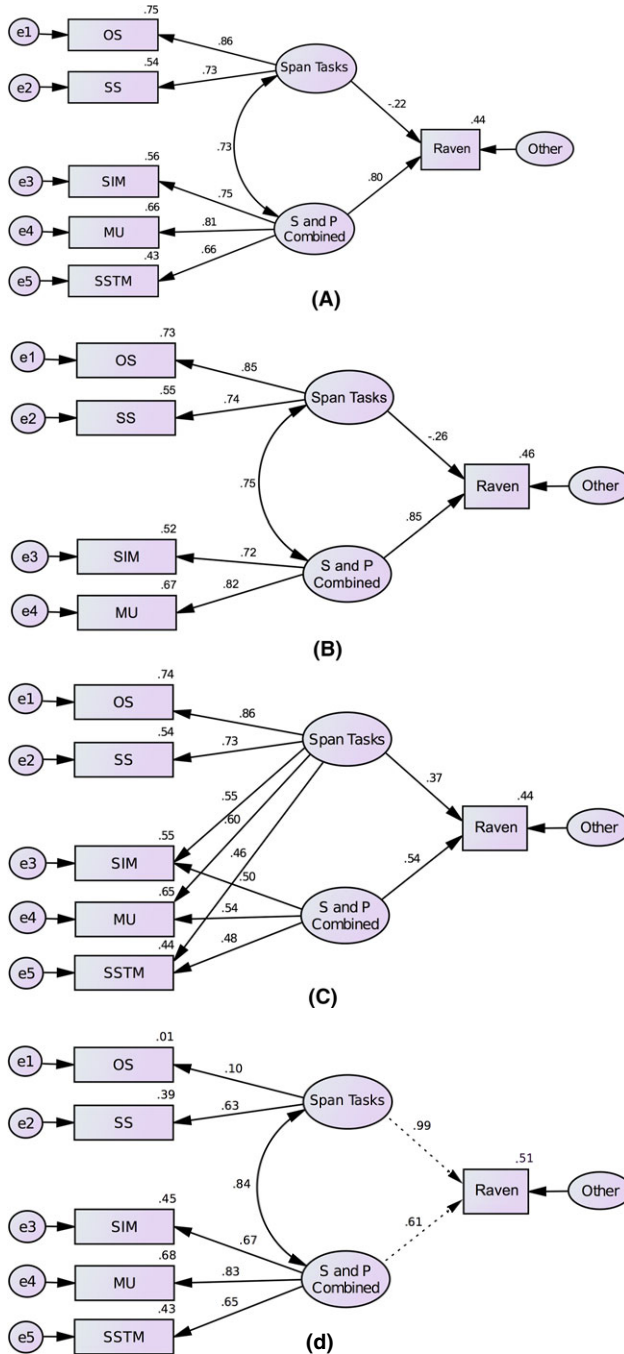
These results suggest that Raven's matrices are better predicted by the construct in which *storage* and *processing* are combined ($r = .64$, corresponding to 41% of the shared variance, instead of $r = .36$ when separated). The combined construct can be assessed in this study by our SIMON span task, a memory updating task, and a simple span task. To make sure that the absence of SSTM would not weaken the predictions too much (since it is the only task that is mainly spatial), we tested a similar model that did not include the SSTM task (Fig. 5b). In this case, we observed higher regression weights for Raven's matrices (.85 instead of .80, $-.26$ instead of $-.22$, and a percentage of shared variance with Raven's matrices of 43% instead of 41%; $\chi^2(3) = 1.22$, $p = .75$). Fig. 5c shows that when a factor already reflects all of the span tasks, a second factor reflecting only the tasks for which *storage* and *processing* are assumed to be combined still account independently for Raven performance. The subtraction of the less restrictive model (Fig. 5a) from the more restrictive model (Fig. 5d) showed that forcing the two paths toward Raven's matrices to be equal significantly reduced the model's ability to fit the data ($\chi^2(9 - 7 = 2) = 54.95 - 2.85 = 52.1$), indicating that the .80 loading in Model A can be considered significantly greater than $-.22$.

5. Discussion

Many complex mental activities still fit our rather low memory capacity. One observation is that challenging concepts (those of experts or those transmitted by culture) have been developed in the simplest possible way to become "simplex" (Berthoz, 2012). Pedagogy is probably a quest for simplicity (for instance, the division algorithms first taught in universities are now taught in elementary school), and language could be under pressure to be compressible (Smith, Tamariz, & Kirby, 2013). STM capacity has been constant for a hundred years (Gignac, 2015), and quite low, and maybe more elaborated

Fig. 5. Path models from confirmatory factor analysis with (a) and without (b) the SSTM task. For comparison, a third model (c) used a factor reflecting all of the span tasks versus another factor reflecting only the three tasks in which *storage* and *processing* were combined. A fourth model (d) further constrained the parameters between Raven and the latent variables to 1 (dotted lines), and a table (e) recapitulates the fit of each model. *Notes.* OS, operation span; SS, sentence span; SIM, chunking span task based on an adaptation of the SIMON[®] game; SSTM, spatial short-term memory; MU, memory updating; S and P, storage and processing; Raven, Raven's advanced progressive matrices. The numbers above the arrows going from a latent variable (oval) to an observed variable (rectangle) represent the standardized estimates of the loadings of each task onto each construct. The paths connecting the latent variables (ovals) to each other represent their correlations. Error variances were not correlated. The numbers above the observed variables are the R^2 values.

reasoning can only occur when the underlying concepts are compressed to fit capacity. However, not all concepts are pre-compressed by previous generations, and sometimes intelligence may require compression to achieve lower memory demands to solve new



problems. We, therefore, tested a new concept of intelligence based on the idea that optimal behavior can be linked to compression of information. Our experimental setup was developed to study this compression process by allowing participants to mentally reorganize the to-be-remembered material to optimize storage in WM. To follow up on our previous conceptualization wherein WM is viewed as producing, on the fly, a maximally compressed representation of regular sequences using chunks (Chekaf et al., 2016; Mathy & Feldman, 2012), we used an algorithmic complexity metric to estimate the compressibility factor. Taken together, our results suggest that the existence of opportunities for compressing the memorandum enhance the recall process (for instance, when fewer colors are used to build a sequence, resulting in low complexity). More interestingly, we found that having more repetitions of colors statistically interacted with sequence length, which indicates that the compressibility factor applies the best to long sequences. Furthermore, although length had a detrimental effect on recall, the complexity effect was found to be stronger than the length effect.

Regarding the relationship between compression and intelligence, the capacity to compress information (estimated by an individual's inflection point along the complexity axis of memory performance) was found to correlate with the performance on Raven's matrices. This result indicates that participants who have a greater ability to compress regular sequences of colors tend to obtain higher scores on Raven's matrices. This correlation is comparable to the one obtained from a composite measure of WM capacity (using the WMCB). However, an exploratory analysis showed that our span task saturated two principal factors in a way very similar to that found for other tasks where processing is also devoted to storage (involving updating or free recoding of spatial information in the absence of a concurrent task), unlike the complex span tasks generally used to estimate Gf (remember that in complex span tasks, processing of the concurrent task is separate from storage in the main task, so it cannot support storage). A confirmatory factor analysis indicated greater predictability for the Raven test with a latent variable in which the storage and processing components were combined, in opposition to a latent variable representing complex span tasks in which storage and processing were separated. This latent variable shared 41% of the variance with Raven's matrices, which seems quite good compared to the 50% obtained by Kane et al. (2005), who used 14 different data sets (on more than 3,000 participants) to estimate the variance shared between the WM-capacity and general-intelligence constructs (see also Ackerman, Beier, & Boyle, 2005, for a contrasting point of view). Our results confirm previous findings that memory updating can be a good predictor of Gf (e.g., Friedman et al., 2006; Salthouse, 2014; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009). More important, we think that the updating task could be associated with a SIMON span task under the same construct. Even though they do not seem to have much in common, processing (either updating or compressing) is dedicated to storage in both of these tasks.

These findings are largely consistent with previous studies suggesting that prediction of Gf by simple spans can reach that of complex spans by increasing list-lengths (Unsworth & Engle, 2007). Increasing digit list-lengths, for instance, might require the participant to further process the sequence to optimize capacity. If the digit storage capacity is

four independent digits (as in the brief visual presentations originally studied by Sperling, 1960), the only way to increase capacity is to recode or group together a few items. Furthermore, the participant's linguistic experience with digits (Jones & Macken, 2015) can presumably account for some additional form of compression due to a greater linguistic experience (see Christiansen & Chater, 2016, who develop the connected idea that the language system might eagerly compress linguistic input), and this observation plausibly supports our conclusion since digit span has strong links to intelligence.

This study allows us to conclude that processing and storage should be examined together whenever processing is fully devoted to the stored items, and we believe that storage and processing must function together whenever there is information to be compressed in the memorandum. Thus, the ability to compress information in span tasks seems to be a good candidate for accounting for intelligence, along with other WM tasks (i.e., simple updating and short-term memory-span tasks) in which storage and processing also function together. This is in line with Unsworth et al. (2009), who argued that processing and storage should be examined together because WM is capable of processing and storing information simultaneously.

Acknowledgments

This research was supported in part by a grant from the Région de Franche-Comté AAP2013 awarded to Fabien Mathy and Mustapha Chekaf. We are also grateful to Nelson Cowan, Yvonnick Noël, and to the Attention & Working Memory Laboratory at Georgia Tech for helpful discussions, and particularly to Tyler Harrisson, who suggested one structural equation model. Authors' contribution: FM initiated the study and formulated the hypotheses; MC and FM conceived and designed the experiments; MC performed the experiments; MC, NG, and FM analyzed the data; NG computed the algorithmic complexity and other measures of complexity; MC, NG, AG, and FM wrote the paper.

Notes

1. Here, we focus on a lossless compression process (i.e., the original data can be entirely reconstructed from the compressed data) and not on lossy compression process (that allows achieving a more substantial reduction in data; however, the original data may be altered).
2. In comparison to MDL, it is fair to admit that ACSS can also lead to longer lengths than the original string, but this is especially true because a program would systematically require a series of `print...end` instructions that mostly impact short strings. However, the comparisons of different short strings is more reliable because the detected regularities in any sequence are systematically recoded using ACSS, contrary to MDL, which might leave some regularity undetected when it does not shorten the original string.

3. Although there are new guidelines for sample size requirements (Wolf, Harrington, Clark, & Miller, 2013), we simply followed the rule-of-thumb of a minimum sample size of 100 participants (those administered with the Raven's matrices) for running structural equation models, including six observed variables.
4. Because repetitions occur within sequences, the proportion of correctly recalled items was not computed, as it unfortunately involves complex scoring methods based on alignment algorithms (Mathy & Varré, 2013), not yet deemed to be standardized scoring procedures.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30.
- Baum, E. B. (2004). *What is thought?* Cambridge, MA: MIT Press.
- Berthoz, A. (2012). *Simplicity: Simplifying principles for a complex world* (G. Weiss, trans.). Yale, CT: Yale University Press.
- 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.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brugger, P., Monsch, A. U., & Johnson, S. A. (1996). Repetitive behavior and repetition avoidance: The role of the right hemisphere. *Journal of Psychiatry and Neuroscience*, *21*, 53.
- Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science. *Trends in Cognitive Sciences*, *7*(1), 19–22.
- Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96–107.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, E62.
- Conway, A. R., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology: Interdisciplinary and Applied*, *44*, 319–327.
- Delahaye, J.-P., & Zenil, H. (2012). Numerical evaluation of algorithmic complexity for short strings: A glance into the innermost structure of randomness. *Applied Mathematics and Computation*, *219*(1), 63–77.
- Dieguez, S., Wagner-Egger, P., & Gauvrit, N. (2015). Nothing happens by accident, or does it? A low prior for randomness does not explain belief in conspiracy theories. *Psychological Science*, *26*, 1762–1770.
- Fass, D., & Feldman, J. (2002). Categorization under complexity: A unified MDL account of human learning of regular and irregular categories. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing* (vol. 15, pp. 34–35). Cambridge, MA: MIT Press.
- Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, *37*, 1565–1578.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.

- Gauvrit, N., Singmann, H., Soler-Toscano, F., & Zenil, H. (2015). Algorithmic complexity for psychology: a user-friendly implementation of the coding theorem method. *Behavior Research Methods*, *48*, 314–329.
- Gauvrit, N., Soler-Toscano, F., & Guida, A. (2017). A preference for some types of complexity comment on “perceived beauty of random texture patterns: A preference for complexity.” *Acta Psychologica*, *174*, 48–53.
- Gauvrit, N., Soler-Toscano, F., & Zenil, H. (2014). Natural scene statistics mediate the perception of image complexity. *Visual Cognition*, *22*(8), 1084–1091.
- Gendle, M. H., & Ransom, M. R. (2009). Use of the electronic game SIMON[®] as a measure of working memory span in college age adults. *Journal of Behavioral and Neuroscience Research*, *4*, 1–7.
- Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence*, *48*, 85–95.
- Hutter, M. (2005). *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. New York: Springer.
- Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*, 1–13.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71.
- Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span. *Memory & Cognition*, *32*(6), 956–964.
- Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254.
- Lewandowsky, S., Oberauer, K., Yang, L. X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*, 571–585.
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York: Springer Verlag.
- Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*, 239.
- Manschreck, T. C., Maher, B. A., & Ader, D. N. (1981). Formal thought disorder, the type-token ratio and disturbed voluntary motor movement in Schizophrenia. *The British Journal of Psychiatry*, *139*, 7–15.
- 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., & Varré, J. S. (2013). Retention-error patterns in complex alphanumeric serial-recall tasks. *Memory*, *21*, 945–968.
- 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.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. H. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York: Oxford University Press.
- Pothos, E. M. (2010). An entropy model for artificial grammar learning. *Frontiers in Psychology*, *1*, 16.
- Raven, J. C. (1962). *Advanced progressive matrices: Sets I and II*. London: HK Lewis.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465–471.
- Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: A MDL-based cognitive model of inductive learning. *Cognitive Science*, *35*, 1352–1389.
- Salthouse, T. A. (2014). Relations between running memory and fluid intelligence. *Intelligence*, *43*, 1–7.
- Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1089.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(379–423), 623–656.

- Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1348–1353). Austin, TX: Cognitive Science Society.
- Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2013). Correspondence and independence of numerical evaluations of algorithmic information measures. *Computability*, 2(2), 125–140.
- Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2014). Calculating kolmogorov complexity from the output frequency distributions of small turing machines. *PLoS ONE*, 9(5), e96223.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74 (1–29).
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038–1066.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, 17, 635–654.
- Wang, T., Ren, X., Li, X., & Schweizer, K. (2015). The modeling of temporary storage and its effect on fluid intelligence: Evidence from both brown–peterson and complex span tasks. *Intelligence*, 49, 84–93.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73, 913–934.