Chunking in Working Memory and its Relationship to Intelligence

Mustapha Chekaf (mchekaf@univ-fcomte.fr)
Département de Psychologie EA 3188, Université de Franche-Comté, 30, rue Mégevand
25030 Besançon Cedex, France

Nicolas Gauvrit (ngauvrit@me.com)
CHArt Lab (PARIS-reasoning), École Pratique des hautes Études, 4-14 rue Ferrus
75014 Paris, France

Alessandro Guida (Alessandro.Guida@univ-rennes2.fr)
CRPCC, EA 1285, Université Rennes 2, Place du Recteur Henri Le Moal
35043 Rennes Cedex, France

Fabien Mathy (Fabien.Mathy@unice.fr)
BCL, CNRS, UMR 7320, Université Nice Sophia Antipolis, 24, avenue des diables bleus
06357 Nice Cedex 4, France

Abstract

Short-term memory and working memory are two distinct concepts that have been measured in simple and complex span tasks respectively. A new span task was designed to manipulate a chunking factor while using a procedure similar to simple span tasks. This span task allowed studying the interaction between storage and processing in working memory, when processing is fully dedicated to optimizing storage. The main hypothesis was that the storage × processing interaction that can be induced by the chunking factor is an excellent indicator of intelligence because both working memory and intelligence depend on optimizing storage. Two experiments used an adaptation of the SIMON® game in which chunking opportunities were estimated using an algorithmic complexity metric. The results show that the metric can be used to predict memory performance and that intelligence is well predicted by the new chunking span task in comparison to other simple and complex span tasks.

Keywords: working memory; span tasks; chunking; information complexity; fluid intelligence

Introduction

The present study is concerned with determining the limits of the short-term memory (STM) span, measured by the length of the longest sequence of items that can be recalled over brief periods of time. One issue when measuring individuals' memory spans is that they are inevitably related to other processes that might inflate their measures, such as information reorganization into chunks (e.g., Cowan, 2001, Cowan, Rouder, Blume & Saults, 2012; Feigenson & Halberda, 2008; Mathy & Feldman, 2012; Miller, 1956) and long-term memory storage (e.g., Ericsson & Kintsch, 1995; Gobet & Simon, 1996; Guida, Gobet, Tardieu, & Nicolas, 2012). We aim to investigate how information reorganization through chunking can be used to optimize immediate recall by developing a new simple span task based on chunking. The new task is based on a measure of complexity that allows one to capture the sum of information that can be grouped to form chunks. In this paper, we examine how this new span measure relates to intelligence and other span tasks measures.

Span Tasks Taxonomy

Simple span tasks traditionally require retaining a series of items (digits, words, pictures), whereas in complex span tasks, participants have to maintain the to-be-recalled material while continuously performing concurrent tasks. Therefore, it has been assumed that short-term memory and working memory refer to the storage and the storage + processing of information respectively. Complex spans have been reported to be better predictors of complex activities and fluid intelligence than simple spans (Unsworth & Engle, 2007a, 2007b), and particularly for Raven's Advanced Progressive Matrices (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). However, simple span tasks are still used in several intelligence tests (such as the Weschler's) since their use with patients in diverse medical contexts is easy, the instructions are simple and the subtests can be done without the need of a computer. Interestingly, Unsworth and Engle (2007a) recently showed that increasing list-lengths could increase the prediction of fluid intelligence in simple spans. These results indicated that simple spans considered to be only "storage tasks" can be viewed as "storage + processing" tasks as well.

Our idea was to devise a memory span task in which both storage and processing could be measured simultaneously and independently, and we argue that this can be done by inducing a chunking process. The main contribution of the present study is to allow independent manipulation of processing and storage, sliding across the [(storage) ... (storage + processing)] continuum, and to investigate the contribution of the processing component to the optimization of storage capacity. A second aim was to measure the relationship between this hypothesized optimization process and fluid intelligence.

Chunking Span Tasks

Several studies have studied the formation of chunks in immediate memory when encouraging chunking and while avoiding long-term learning effects (Bor et al., 2004, 2003; Bor & Owen, 2007; Mathy & Feldman, 2012; Mathy & Varré, 2013). The present study continues this line of research in order to show that participants exposed to simple sequences of colors show higher recall for more regular sequences without particular relation to prior knowledge in long-term memory. Our new task is based on the framework of SIMON®, a classic memory game from the 80s that consists of immediately reproducing sequences of colors. The device lights up colored buttons at random and increases the number of colors by adding a supplementary color at the end of the previous sequence whenever the reproduction by the player is correct. This task has interesting properties as it is resistant to practice effects, habituation, and proactive interference across trials (Gendle & Ransom, 2006). There were two important differences between the original game and the present adaptation. First, a given chosen sequence was not presented progressively but entirely in a single presentation. For instance, instead of being presented with a "1) blue, 2) blue-red, 3) blue-red-red, etc.", that is, three series of the same increasing sequence until a mistake was made, the participant in this case would be given a blue-red-red sequence from the outset. If correct, a new sequence was given, possibly using a different complete length, so there was no sequence of increasing length that could have favored a long-term memory process. Second, no sounds were associated with any of the colors.

Complexity for Short Strings

To estimate the chunking opportunities of the sequences of colors, a compressibility metric was sought to provide an estimation of any possible grouping process. More complexity means less chunking opportunities. Less complexity means that a sequence can be re-encoded for optimizing storage and in this case, our idea is that processing takes precedence over storage. A major difficulty one encounters in this type of study is due to the apparent lack of a normalized measure of compressibility—or complexity. Some formal measures such as entropy are actually widely used as proxy for complexity, but they have come under harsh criticism (Gauvrit, Zenil, Delahaye, & Soler-Toscano, 2014). For instance, entropy only depends on the relative frequencies of the different outcomes. Thus, according to entropy, the two strings "*red-blue-red-blue-red-blue-red-blue*" and "*red-blue-blue-red-blue-red-red-blue*" are equally complex since the two colors appear in the same proportion in each sequence. The fact that the first one can be compressed as "4 times *red-blue*" is not detected by entropy. Our compressibility metric is based on algorithmic complexity, which formally is defined as the length of the shortest program that outputs a string (Li & Vitányi, 2009). Contrary to long strings, the algorithmic complexity of short strings (3-50 symbols or values) could not be estimated before recent breakthroughs (Delahaye & Zenil, 2012; So1er-Toscano, Zenil, Delahaye, & Gauvrit, 2013, 2014), thanks to which it is now possible to obtain a reliable estimation of the algorithmic complexity of short strings (3-50 symbols or values).

The algorithmic probability $m(s)$ of a string s is defined as the probability that a randomly chosen deterministic program running on a Universal Turing Machine will produce s and halt. This probability is related to algorithmic complexity by way of the algorithmic coding theorem which states that $K(s) \sim -log_2(m(s))$, where $K(s)$ is the algorithmic complexity of s. Instead of choosing random programs on a fixed Universal Turing Machine, one can equivalently choose a random Turing Machine and have it run on a blank tape. This has been done on huge samples of Turing machines (more than 10 billions Turing Machines), and led to a distribution d of strings, approximating m. The algorithmic complexity for short strings of a string s, acss(s) is defined as $-log_2(d)$, an approximation of $K(s)$ by use of the coding theorem.

The idea is not that the human brain operates as Turing machines, but in fact, this method is used here to provide approximations to capture any kind of regularities in a string. Algorithmic complexity is, in a way, the normative ultimate measure of compressibility or "chunkability". For example, Table 1 shows length and complexity of a random sample of sequences used in our chunking span tasks and presented to the participants, after being coded into sequences of colors.

Table 1: Examples of sequences; Note: each digit codes for a specific color, for example, "31131331" codes for "*red-blue-blue-red-blue-red-red-blue*"

Sequence	Length	Complexity
2223332	7	21.4476
2232113	7	22.5040
12121212	8	22.7576
31131331	8	24.8765
424242244	9	26.7262

Relationship Between Storage × Processing and Intelligence

Bor and colleagues (Bor, et al., 2004; Bor, Duncan, et al., 2003; Bor & Owen, 2007) introduced systematic regularities that encouraged the participants to chunk redundancies in list of digits. More formally, chunking was induced in our task by manipulating the algorithmic complexity of the to-be-remembered series of colors, which allowed varying gradually the probability to form a chunk in working memory.

It is assumed that the most complex sequences cannot be easily reorganized and as such they reduce processing opportunities and mainly involve *storage*. Conversely less complex sequences are assumed to favor the occurrence of chunking via reorganization of the material and should thus involve *storage* × *processing*. This interaction aims to identify situations where an individual having low storage and high processing capacities could obtain a span similar to someone having high storage and low processing capacities.

Experiment 1 only aimed at studying the *storage* × *processing* capacity and verified its relationships to other span tasks and IQ. Experiment 2 used two conditions enabling us to hypothesize a better separation between the *storage* and *storage* × *processing* capacities, and their relationships to other span tasks and IQ.

Experiment 1 was very liberal with random sequences of colors, which lead us to develop a specific estimation of memory capacity. Experiment 2 on the contrary used similar sequences across participants that allowed us to use a more standard scoring method for computing a memory span.

Experiment 1

Method

Participants and Procedure

The tests were administered to 183 students (M_{age} = 21; SD = 2.8) in the following order: the Chunking span task (SIMON), the Working Memory Capacity Battery (WMCB) (Lewandowsky, Oberauer, Yang & Ecker, 2010) and Raven's APM (Raven, 1962) (set #2, 40 minutes). The WMCB includes four tasks: a memory updating task (MU), two complex span tasks (operation and sentence span, OS and SS), and a spatial simple span task (spatial short-term memory, SSTM).

Chunking Span Task

Fifty random to-be-memorized sequences of colored squares appearing one after the other was displayed (see Figure 1). In the recall phase, four colored buttons were displayed and participants could click on them to recall the sequence they had memorized.

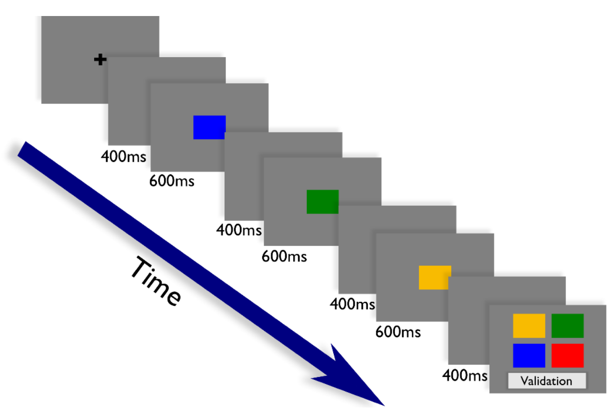

Figure 1: Example of a sequence of three colors of the Chunking Span Task.

Results

Effect of Complexity

The relationship between complexity measure and sequence length seems quite obvious. However, in order to assess and compare the respective impacts of complexity and list-length we used a logistic regression. A stepwise forward model selection based on BIC criterion suggested dropping the interaction term. This model showed a significant negative effect of complexity ($z(9147) = -23.84$, $p < .001$, st. coef. = -5.7 [coef. -.69]) as shown in Figure 2, and although length had a detrimental effect on recall ($z(9147) = 16.27$, $p < .001$, st. coef. = 3.74 [coef. 1.46]), this effect was more than compensated by the detrimental effect of complexity, meaning that long simple strings were easier to recall than shorter but more complex strings. In other words, the effect of complexity was stronger than the effect of length.

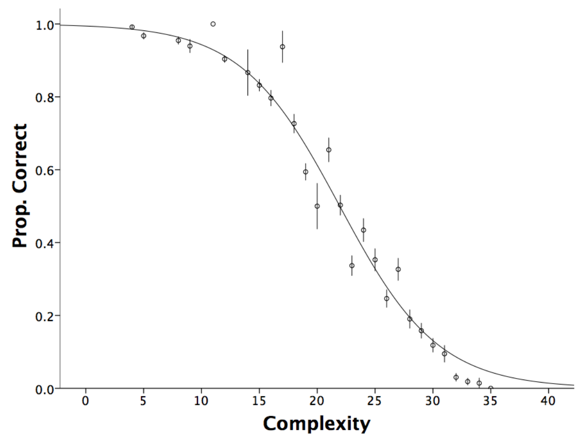

Figure 2: Proportion correct as a function of complexity in Experiment 1. Note: Error bars are +/- 1 SE.

Correlations and factor analysis

Table 2 shows the correlations between measures aggregated by participants, including the global WM score for the entire WMCB, the Raven's APM (RAVEN), and the Simon span task (SIMON). The correlation matrix shows that in terms of prediction of the Raven's score, the Simon's score is comparable to the composite WM score produced by the WMCB (this range of correlations corresponds to that found in the literature).

Table 2: Correlation matrix for Experiment 1. Note: ** $p < .001$.

	SIMON	WM
RAVEN	.428**	.437**
SIMON		.531**

One important aspect to recall is that SIMON, MU, and SSTM each allow the stored items to be processed while both OS and SS are standard complex span tasks that separate processing and storage. One hypothesis was that performance on the SIMON span task should better correlate with MU and also with SSTM. A second prediction was based on the idea that the tasks in which the stored items are fully processed (i.e., involving storage × processing) would better predict the average Raven's score. The correlation between MU and the Raven was effectively the highest (r = .572), and the Simon was the second task to better correlate with the Raven. The Simon also best correlated with both MU and SSTM (two tasks in which processing is effectively dedicated to storage).

We conducted a principal component analysis to extract two factors, which were expected to separate a storage component from a processing component. The two components accounted for 40% and 30% of the variance respectively (the respective eigenvalues being 2.4 and 1.8, instead of 3.3 and .89 for the unrotated initial solution). We interpreted the two factors as clearly separating the complex span tasks (in which processing is estimated alone, while processing is saturated) and the tasks in which processing was dedicated to storage, but it is still difficult to see how the processing and storage components are separated in these analyses by the respective factors. The data were then submitted to a confirmatory factor analysis using IBM SPSS AMOS 21. A latent variable representing a construct in which storage and processing are separated and another latent variable representing a construct in which both processes interact (the processing component) were sufficient to accommodate performance. The fit of the model is shown in Figure 3 ($\chi2(7)$ = 2.82, p = .90; CFI, comparative fit index = 1.0; RMSEA, root mean squared of approximation = 0.0; RMR, root-mean square residual = .063). These results, confirmed by a comparison of correlation coefficients (z = 15.7, p < .001), showed that the Raven's scores are better predicted by the construct in which storage and processing are combined (r = .64, corresponding to 41% of shared variance, instead of r = .36 when separated), a construct that can be reflected in the present study by our chunking span task, a memory updating task, and a spatial simple span task.

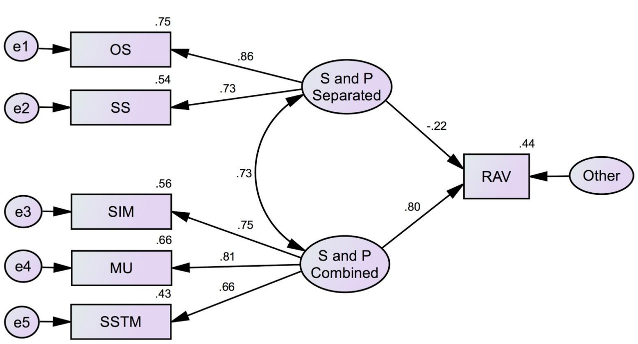

Figure 3: Path model for factor analysis from Exp. 1. Correlations and loadings are std. estimates. OS, operation span; SS, sentence span; SIM, chunking span task; MU, memory updating; S and P, storage and processing; RAV, Raven's APM.

Experiment 2

Method

Participants and Procedure

The tests were administered to 107 students (M_{age} = 22.9, SD = 5.9) in the following order: the Simon chunking span task (SIM), the digit simple-span subtests of the WAIS-IV (Wechsler, 2008): the Digit Span Forward (DSF) which requires recalling a series of digits in correct order, the Digit Span Backward (DSB) which requires recalling a series of digits in reverse order, and the Digit Span Sequencing (DSS) which requires recalling a series of digits in ascending order, and finally the Raven (set #2, 40 minutes; N = 95 because the Raven was optional for getting extra course credits).

Chunking Span Task

Procedure and scoring of the span followed that of the WAIS: the length of the presented sequences progressively increased, starting with length two, then three, etc. The longest span attained at least once was considered as the subject's span. Each participant was administered two complexity conditions. The Simple condition was conducive to inducing a chunking process, while the Complex condition allowed less chunking opportunities and as such was considered as mostly soliciting the storage component.

Results

Effect of Complexity

The logistic regression (see Figure 4) showed a detrimental effect of complexity on recall ($z(2651)$ = 9.77, p < .001, st. coef = -6.41 [non standardized: -.95]) and also an effect of length (z = 6.273, p < .001, st. coef= 3.94 [non standardized: 2.01948]).

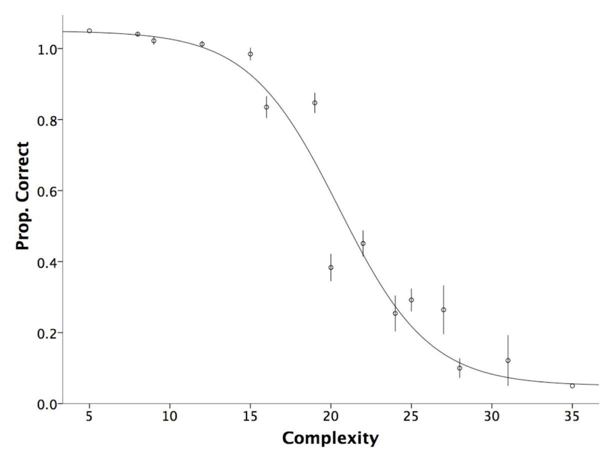

Figure 4: Proportion correct as a function of complexity in Experiment 2. Note: Error bars are +/- 1 SE.

Correlations and factor analysis

Despite the moderate difference between the two mean spans observed between the simple and complex conditions, Table 3 shows that these two conditions highly correlated

(r = .42), in comparison with other variables. Similarly, DSF and DSB shared the greatest percentage of variance (r = .47, p < .001), as well as DSB and DSS (r = .48, p < .001). Thus the digit spans showed high mutual correlation, but none of the digit span tasks correlated more with either Simon simple or Simon Complex than the two together. One possibility is that the participants could still chunk many of the most complex sequences, making the two Simon conditions akin, and accounting for the slight difference of 1.25 colors reported above. The possibility that participants chunk the less compressible sequences does not contradict compressibility theories because the estimate of the compressibility of a string is an upper bound (meaning that there can always be a way to compress a string more than it is expected). Regarding correlations with the Raven, the highest correlation was found with DSB, but the multicollinearity of the data makes interpretation of the pairwise correlations difficult.

Table 2: Correlation matrix for Experiment 2. Note: SIMPL, COMPL, mean span in simple and complex conditions; DSF, DSB, DSS, Digit Span Forward, Backward, Sequencing; Note: ** p < .001.

	Compl	DSF	DSB	DSS	RAV
Simpl	.422**	.294**	.337**	.157**	.413**
Compl		.229**	.353**	.310**	.385**
DSF			.473**	.273**	.290**
DSB				.476**	.446**
DSS					.297**

Principal component analysis was used to explore our data and to extract two factors (which were expected to separate the chunking span tasks and the WM span task). The two factors clearly separated the digit span tasks and the chunking span tasks. It is worth noting that the Raven loaded with the chunking span tasks.

The data were submitted to a confirmatory factor analysis using IBM SPSS AMOS 21 in order to test the prediction that tasks allowing the processing and storage components to fully function together in association to optimize storage are better predictors of general intelligence than the STM span tasks of the WAIS. A latent variable representing a chunking construct (derived from the Simon span tasks) and another latent variable representing a simpler STM construct (derived from the digit span tasks of the WAIS) were sufficient to accommodate performance. The fit of the model shown in Figure 5 was excellent ($\chi 2(7)$ = 3.2, p = .87; CFI, comparative fit index = 1.0; RMSEA, root mean squared of approximation = 0.0; RMR, root-mean square residual = .049; AIC and BIC criterions were both the lowest in comparison to a saturated model with all the variables correlated with one another and an independence model with all the variables uncorrelated). These results, confirmed by a comparison of correlation coefficients (z = 1.82, p = 0.03), showed that the Raven's scores are best predicted by the Chunking latent variable, a construct that can be reflected in the present study by the two chunking span tasks.

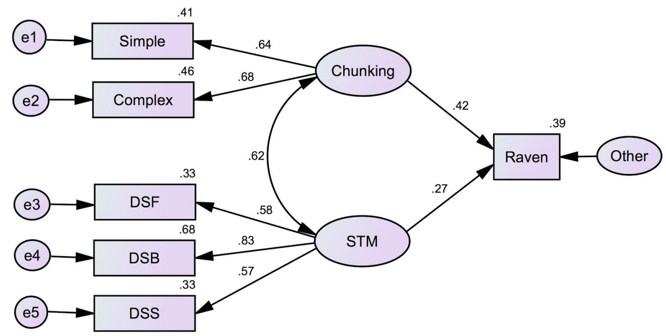

Figure 5: Path models for confirmatory factor analysis from Exp. 2. Correlations loadings are std. estimates. Note: Simple, Complex, simple and complex conditions; DSF, DSB and DSS, Digit Span Forward, Backward and Sequencing.

General Discussion

In two experiments, we found that span tasks involving a chunking process were structurally closer to the performance on the Raven's than any complex span task of the WMCB and any of the simple span tasks of the WAIS. The present study shows that simple span tasks can effectively compete with complex span tasks, and this was achieved here by prompting the creation of chunks in immediate memory while avoiding a long-term learning effect.

Therefore, it seems likely that the span tasks better correlate with higher cognitive processes when they prompt reorganization of information. The present study concludes that processing and storage should be examined together when processing is fully dedicated to the stored items, and we believe that the interaction between storage and processing that best represents a chunking process in immediate memory can provide a true index of WM capacity. This is in line with Unsworth, Redick, Heitz, Broadway & Engle (2009) who argue that processing and storage should be examined together because WM is capable of processing and storing information simultaneously.

Conclusion

The rationale of the present study was that sequences of colors of the Simon game contain regularities that can be mathematized to estimate a chunking process, and that the quantity of chunking induced in a to-be-recalled sequence can represent the processing demand. The chunking span task allows the processing and storage components to fully interact to optimize storage. Although it is not commonly accepted in the literature that span tasks can take benefit from favoring the processing of the stored items (which explains the plethora of complex span tasks in the literature), the chunking span task was found a reliable predictor of general intelligence in comparison to other simple or complex span tasks.

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 grateful to Caroline Jacquin for her assistance in data collection in Exp. 2. Authors' contribution: FM initiated the study and formulated the hypotheses; MC and FM conceived and designed the experiments; MC run the experiments; MC, NG, and FM analyzed the data; NG computed the algorithmic complexity; MC, NG, AG, and FM wrote the paper.

References

Bor, D., Cumming, N., Scott, C. E., & Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *European Journal of Neuroscience*, *19*(12), 3365–3370.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*(2), 361–367.

Bor, D., & Owen, A. M. (2007). A common prefrontal–parietal network for mnemonic and mathematical recoding strategies within working memory. *Cerebral Cortex*, *17*(4), 778–786.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide, *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87-114.

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: what does it take to make them work? *Psychological review*, *119*(3), 480-499.

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.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*(29), 9926–9930.

Gauvrit, N., Zenil, H., Delahaye, J.-P., & Soler-Toscano, F. (2014). Algorithmic complexity for short binary strings applied to psychology: a primer. *Behavior Research Methods*, *46*(3), 732-744.

Gendle, M. H., & Ransom, M. R. (2006). 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.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*(2), 159–163.

Guida, A., Gobet, F., & Nicolas, S. (2013). Functional cerebral reorganization: a signature of expertise? Reexamining Guida, Gobet, Tardieu, and Nicolas'(2012) two-stage framework. *Frontiers in Human Neuroscience*, *7*.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*(2), 571–585.

Li, M., & Vitányi, P. M. (2009). *An introduction to Kolmogorov complexity and its applications*. New York, NY: Springer Verlag.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362.

Mathy, F., & Varré, J. S. (2013). Retention-error patterns in complex alphanumeric serial-recall tasks. *Memory, 21*(8), 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*(2), 81.

Raven, J. C. (1962). *Advanced progressive matrices: Sets I and II*. HK Lewis.

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.

Unsworth, N., & Engle, R. W. (2007a). 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*(6), 1038.

Unsworth, N., & Engle, R. W. (2007b). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104.

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*(6), 635–654.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*.