How Working Memory Capacity Constrains the Learning of Relational Concepts

Adam Chuderski (adam.chuderski@uj.edu.pl)
Institute of Philosophy and Institute of Psychology, Jagiellonian University
Grodzka 52, 31-044 Krakow, Poland

Anna Chuderska (ania.chuderska@gmail.com)
Institute of Psychology, Jagiellonian University
Mickiewicza 3, 31-120 Krakow, Poland

Abstract

We investigated the way in which working memory (WM) constrains the learning of relational concepts – categories defined by the way objects are assigned to roles in the structure of an underlying relation, and not by objects' intrinsic features. By applying to a large sample a novel test of concept learning as well as the battery of WM tasks, we found that WM is a strong predictor of the scores on the test, but the WM-learning correlation decreases as the relational complexity of the to-be-learned concepts increases. Such results support those theoretical models of relational learning, which assume that learning of relational concepts (and relations, in general) consumes more WM resources than just the processing of relations which have already been learned.

Introduction

The issue of relational thinking – the humankind's ability to acquire, process, and effectively use mental representations of relations – has huge importance in cognitive science (Gentner & Kurtz, 2005; Halford, Wilson, & Phillips, 1998; Hummel & Holyoak, 2003). A relation can be described as an ordered list (a structure, a predicate) of well-defined roles and objects that fulfill them. The key aspect of relations consists of the fact that understanding of them as well as inferring from them depends primarily on the way objects are assigned to roles in the relation's structure, and not necessarily on objects' intrinsic features. Relational representations constitute the core of human complex cognition, including abstraction, reasoning, analogy making, creativity, and language (Halford, Wilson, & Phillips, 2010).

The extension of an *n*-ary relation (where *n* is a number of roles in a relation; its arity) is a subset of Cartesian product of *n* sets, which includes all lists of objects (*n*-tuples) that can fulfill roles in that relation (i.e., an object from the first set in a tuple fulfills the first role, etc.; Halford et al., 1998). So, each relation can be treated as a *relational category/ concept* (Gentner & Kurtz, 2005). Unlike so-called *entity categories*, that is, categories formed by objects due to their perceptual or/and internal (e.g., genetic) similarity (e.g., natural kinds), relational concepts in the first place organize entity categories (or lower level relational categories), and so their exemplars may drastically vary featurally. For example, the instances of the relational concept of *barrier* will include: a wall, a river, a person, but also an insult, and loss of support (ibidem). Relational concepts constitute the main part of culture, science, and technology.

A key goal of cognitive science is to understand what relational concepts are, how they are acquired in childhood and adulthood, and how they are used in relational thinking. Consequently, the present paper aims to deal with one specific problem in this domain: it investigates in what way the constraints of human cognitive architecture, particularly its working memory capacity (WMC), influence the learning of relational concepts (from here on, the process/ability referred to as *relational learning*).

Working memory and relational learning

Computational models of relational thinking (e.g., Chuderski, Andrelczyk, & Smolen, 2013; Doumas, Hummel, & Sandhofer, 2008; Halford et al., 2010; Hummel & Holyoak, 2003) as well as psychometric studies on reasoning and analogy making (e.g., Martinez et al., 2011) suggest that processing relations is grounded in working memory (WM). WM is a neurocognitive mechanism responsible for maintenance of a limited, but crucial for the current task/goal, amount of information, in an active and easily available state (Cowan, 2001). It thus allows for flexible manipulation of that information (Hummel & Holyoak, 2003; Oberauer, Süß, Wilhelm, & Sander, 2007), including binding of relational roles and corresponding objects, which is a necessary process for a relational representation to be constructed. People can hold in their WM up to, on average, as few as three or four chunks of information (Cowan, 2001; Luck & Vogel, 1997) and, probably, the similar number of bindings (Chuderski et al., 2013; Oberauer et al., 2007), though these values vary among individuals (approx. from 1 to 6). This clearly corresponds to the fact that accuracy of processing relations sharply decreases with increasing arity of relations, and few participants can cope with relations more complex than quaternary ones (Halford et al., 2010).

An interesting research question pertains to a problem of whether similar influence of WMC, as in abovementioned case of processing relations (e.g., during analogical mapping or inference), also takes place in case of relational learning, when people have to discover an (abstract) relation between related objects and construct a mental representation of the relational concept referring to that category.

A widely used paradigm of relational concept learning was proposed by Shepard, Hovland, and Jenkins (1961). They presented to participants series of eight three-feature geometric figures, each of which could take one out of two values on each featural dimension (shape, size, color), and needed to be classified as belonging or not to a category defined by an arbitrary rule of propositional logic (henceforth named a *Boolean concept*). Each decision was followed by a feedback information on whether an object was categorized correctly or not. Participants improved on that classification task, thus learning to some extent a hidden Boolean concept. The most important result suggested that the accuracy of categorization decreased as the number of features relevant for a concept increased from one to three. This observation was generalized beyond Boolean concepts domain by Halford et al. (1998), who defined relational complexity in terms of the number of entities (variables or dimensions of a relation, that is, its arity) that must be related in parallel, because their decomposition into a set of less complex relations would lead to the loss of the relation's meaning.

Moreover, accuracy of categorization decreased with increasing complexity (expressed by minimal description length, MDL; Feldman, 2000) of a logical rule associated with the three-feature concepts. However, as the MDL approach led to a problem of which logical operators should count as minimal (e.g., if we include exclusive disjunction, then MDL no longer predicts Shepard et al.'s data; see Goodwin & Johnson-Laird, 2011), later this approach was disputed. For example, Feldman (2006) proposed an algebraic complexity metric of discrete-value concept learning difficulty, which depends on the sum of a number of constant values of variables and the number of implications which derive values of one variable from values of another variable, into which a given concept can be decomposed. Kemp, Goodman, and Tenenbaum (2008) adopted this approach to describe relational concepts beyond a Boolean domain. Those approaches nicely predicted observed data. Similarly good fit was obtained by a theory predicting that the number of all possible mental models (iconic-like representations precisely corresponding to the structure of – themselves roughly represented – elements of a situation) which match a rule describing a concept (Goodwin & Johnson-Laird, 2011).

In the present paper, we ask whether the effectiveness of relational learning can be predicted by WMC. Moreover, we test whether the link between those two variables, if any is found, depends on the abovementioned complexity of concepts which are learned. Such a test may be very informative regarding the validity of existing models of relational learning, because, as we will see, some of them seem to yield opposite patterns of predictions on the strength of WMC-relational-learning link in the function of complexity.

Although Halford et al. (1998) have not inferred such predictions directly from their theory (instantiated also in a computational model called STAR), closer inspection of this theory leads to the prediction that the critical value of relational complexity for learning relations should be four dimensions. For example, Halford, Baker, McCredden, & Bain (2005) have shown that accuracy to understand statistical interactions is quite high for two- and three-way interactions, while it radically falls down in the case of four-way ones. As Halford et al. (2010) assume that the same constrains pertain to both processing and acquiring relations (both limits are grounded in the maximal size of a tensor that humans can mentally represent), learning bi- and ternary relations should be relatively easy and not so much constrained by individual WMC, as the mean WMC is about four. In contrast, there should be substantial differences in learning quaternary relations, as people of WMC below four (i.e., one, two, or three) will not be able to learn them fully, while people of WM above that limit (i.e., of four, five or six slots) will have enough capacity to do that. So, the correlation between WMC and relational learning should be the strongest in case of the mean value of WMC.

In contrast, a neurosymbolic model of the discovery of relations proposed by Doumas et al. (2008) assumes that in order to learn a relation, a cognitive system has to represent each role-filler pair as two separate neuronal oscillations, asynchronic, but peaking close in time. This implicates that for learning each dimension of a relation, the system needs two WM chunks, and only after having learned it, both a role and a filler can be compressed into one synchronized oscillation. So, even learning binary relations will consume WMC (i.e., four chunks) of a large part of participants, and their performance on binary relations should be particularly sensitive to individual differences in WMC. Learning ternary (i.e., requiring six WM slots) or quaternary (i.e., occupying eight slots) relations should be difficult for almost everyone's WM, and – if nevertheless effective – will have to rely on mechanisms other than WM (e.g., relational knowledge accretion, compressing relations, etc.).

Interestingly, a recent study by Lewandowsky (2011), who examined correlations between each type of Shepard et al.'s concepts and WMC, has shown that the strength of such a correlation is basically the same in case of unary, binary and ternary concepts of such a kind. This study suggests that a third possibility regarding the pattern of correlations between relational learning and WMC is possible, specifically that WMC influences learning relations of any complexity. However, three disputable aspects of the Lewandowsky's study suggest that more data is needed before a decisive conclusion on WMC-relational learning link can be given.

Firstly, the criterion for a successful learning of Shepard et al.' concepts was that a certain number of correct classifications can be consecutively made by a participant. However, this does not guarantee that he or she really started to represent a relation underlying the concept, because due to a large number of classification trials a complex association, instead of a fully-blown relational representation, may be formed as well. So, in order to prevent such a case, participants should be able to explicitly report a relation to be found, as a necessary criterion for judging that a relational representation has indeed been learned.

Secondly, with the use of Shepard et al.'s concepts, at most ternary relations can be investigated, what does not allow to directly test predictions derived from Halford et al. (1998, 2005, 2010). More complex relations, above and beyond binary features and three dimensions, are needed (e.g., Kemp et al., 2008). Optimally, participants should be required to learn quaternary relations, in which each variable depends on the values of three other variables.

Finally, all existing studies (e.g., Goodwin & Johnson-Laird, 2011; Halford et al., 2005; Kemp et al., 2008; Lewandowsky, 2011) have investigated relational learning defined as the rate of success in acquisition of a to-be-learned concept/rule. However, learning is a dynamic process, and its most important indicator is the increase of knowledge one has gained, and not the total amount of knowledge (including the knowledge possessed before the study started) that one can display. So, the examination of the progress in the discovery of relations, and not only how one can discover them in general, as well as the testing of possible associations between the rate of that progress and WMC, can bring a vital insight about relational learning and its WM mechanisms. To our knowledge, no study so far addressed all aforementioned issues in parallel.

In the remaining part of the paper, we present a direct examination of possible predictions on the link between WMC and relational learning, by applying to a large sample of participants a test that requires discovery of relational concepts differing in complexity. Each discovered concept must aptly describe six presented associated exemplars, while excluding three accompanying counterexamples. We also measured participants' WMC with four versions of a well-established WM measure (a complex span task). We investigated the resulting correlations with the use of confirmatory factor analysis (CFA). Firstly, we correlated relational learning accuracy with WMC, in the function of the complexity of the former. Secondly, we tested the link between the latter and the improvement on learning, that is, when structurally identical relations must be discovered for the second time, but now governing new kind of stimuli. That is, we examined if the transfer of the effects of relational learning would be linked to WMC or not.

A study

Participants

A total of 243 participants (142 women, mean age = 24.3 years, *SD* age = 5.0, range 18 – 45 years) were recruited via publicly accessible social networking websites. Each person was paid 15 euro for their participation in the study. Data from six people were discarded because of theirs failure to provide even one elaborate description in the learning test.

A test of relational concepts discovery

The DREL (Discovery of RELations) paper-and-pencil test consists of two, letter and digit, parts. Each part includes 15 items. Each item consists of six four-symbol strings, which are governed by a to-be-discovered relation, and another three strings, which form counterexamples for that relation, that is, the discovered relation must exclude all three counterexamples. A participant is required to write down a concise and abstract description of a relation that matches six positive exemplar strings. The counterexamples were introduced in order to prevent describing too general relations (e.g., *all strings consist of four symbols*). In each part of the test, there are five binary, five ternary, and five quaternary relations, and item positions for each complexity level with regard to the beginning of the test were balanced (the sequence of levels is: 3 2 4 2 3 4 3 4 2 4 3 2 4 2 3).

In the first part of the test, symbols in each string are two different letters, and a relation governs the place of each letter relative to some number of remaining letters in a string. We assumed that in binary-relation items, the proper relation can be discovered using only pairwise comparisons of letters, so in each step of analysis of a string, a participant needs to maintain in WM only two representations. One example of a binary-relation item requires to discover a relation *the same letters in the middle are different from the same two letters on the extremes*:

OEEO LSSL BVVB

ZKKZ NUUN YAAY

~~RRVV AKAK PPLL~~

Counterexamples prevent people from proposing relations like *there are always two exemplars of one letter and two – of the other letter*. There is only one mental model corresponding to binary relations (in case of this example: *abba*).

In the ternary-relation items, the proper relation can be discovered using comparisons of three letters in parallel, so in each step of analysis three representations have to be maintained in WM. An instance of ternary relation is *one and only letter different from three other identical letters is always placed in the middle* (corresponding models are: *aaba* and *abaa*). In the item presented below, a participant is expected to relate: a pair of two identical letters to another identical letter on the opposite, and both of them to one remaining different letter always placed in the middle:

ZEZZ LLUL NRNN

ASAA JJWJ PBPP

~~OLLL KKKN VVVB~~

In the most difficult, quaternary-relation items, we assumed that all four letters have to be related in one step. An example relation is *the first letter is different from the second one or the third one or both, and the third letter is different from the fourth one* (three corresponding models: *aaba*, *abab*, and *abba*). The complexity of this relation is a result of introducing an inclusive disjunction *x or y or both.* A participant in this example is expected to simultaneously relate the first letter to the second, the first one to the third, and the third one to the fourth:

GGRG NHNH FDDF

BEEB OOXO ACAC

~~FFFF NNNP JJSS~~

The only difference between the first and the second part of the test is that symbols are digits, and relations pertain to their evenness or oddness. However, the abstract structure of relations of corresponding items in both tests is identical. For example, the digit version of aforementioned binary relation would be: *two digits in the middle are both odd or both even, and in the former case two extreme digits are even, while in the latter case two extreme digits are odd*. This part is more difficult, as the crucial feature (evenness/ oddness) is not linked to the appearance of a symbol, while the crucial feature of the letter part (identity/difference) is.

The scoring on the test depended on the abstractness on given descriptions. One point was scored if a described relation was correct and properly abstract (as in the examples), no matter what exact formulation were used by participants. Half point was scored if a description was correct, but it was not abstract enough, instead it was composed of particular subcategories of strings (usually corresponding to possible models), for example, in case of the ternary example, if a description was like *there is either (a) one letter, then another is different, and then two last letters are the same as the first one, or (b) there are two same letters, then another is different, and then the last letter is the same as the two first ones*. No score was given for incorrect descriptions, no matter if they excluded valid instances of strings or included counterexamples. Such a partial scoring resulted in much better reliability of the test (Cronbach's α = .91) than did binary (correct/incorrect) scoring (α = .78). The dependent variables were total scores (in range 0 to 5) on each level of relational complexity, and the corresponding differences between scores on the second and first part of the test (i.e., indices of learning).

Working memory tasks

Four complex span tasks were designed following Conway et al. (2005). In general, a complex span requires memorizing a sequence of a few stimuli, each of them followed by a simple decision task. In the present versions, each task required memorizing three to seven (set size) stimuli, presented for 1.2 s apiece, out of nine possible ones for that task. After two two-stimuli training trials, three trials for each set size (in increasing order) were presented in each complex span task. The letter span task (sometimes called an operation span task) required memorizing letters, while deciding with a mouse button if intermittent simple arithmetical equations (e.g., $2 \times 3 - 1 = 5$?) are correct or not. The digit span consisted of memorizing digits, while checking if letter strings begin and end with the same letter. The spatial span task required memorizing locations of a red square in the 3×3 matrix, while deciding which of two presented bars is larger (the difference was always 25%). In the figural span task, participants were instructed to memorize simple geometric figures, while judging colors to be light (yellow or beige) or dark (brown or navy blue). The dual (decision) task in each WM test aimed to prevent the chunking of stimuli or the extensive use of phonological loop, which could obscure "real" WMC of individuals. The participants were instructed that they should recall as many stimuli as they can (in proper order), but that they should also try to be correct on the decision tasks.

The response procedure in each task consisted of a presentation of as many 3×3 matrices as was a particular set size, in the center of the computer screen, from left to right. Each matrix contained the same set of all nine possible stimuli for a task. A participant was required to point with the mouse those stimuli that were presented in a sequence, in the correct order (from left to right). Only choices that matched both the identity and ordinal position of a stimulus were taken as correct answers. The dependent variable for each complex span task was the proportion of correct choices to all stimuli presented in the task. All complex span tasks displayed high reliability (αs = .85 to .89).

Procedure

The presented study was a part of a larger project testing various cognitive abilities (WM, attention, reasoning), which included 17 computerized tasks applied in one four-hr session, and 5 tests of relational thinking applied in another four-hr session (sessions were administered in a random order), with a 1-hr break between the sessions. Complex span tasks were the 5th, 9th, 13th and 16th tasks in a row applied in the former session, while the DREL test was the first task in the latter session. Half hour was allowed for each part of the DREL test.

Results

Table 1 shows the descriptive statistics and correlations of all dependent variables. No variable deviated from the normal distribution. Correlations ranged from moderate (r = .21) to strong (r = .75).

Task	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.DL2	–									
2.DL3	.46	–								
3.DL4	.36	.66	–							
4.DD2	.42	.45	.44	–						
5.DD3	.27	.46	.47	.65	–					
6.DD4	.23	.43	.52	.56	.75	–				
7.LSPAN	.36	.39	.34	.37	.40	.31	–			
8.NSPAN	.42	.37	.30	.38	.33	.26	.70	–		
9.SSPAN	.21	.32	.23	.34	.34	.25	.57	.51	–	
10.FSPAN	.24	.29	.23	.36	.32	.28	.65	.72	.59	–
Mean	4.46	2.18	1.53	3.07	1.37	0.88	0.69	0.76	0.52	0.62
SD	0.95	0.85	0.99	1.46	0.96	0.96	0.19	0.16	0.18	0.18
Min.	0	0	0	0	0	0	0.05	0.09	0.05	0.13
Max.	5	4	4.5	5	3.5	3	0.99	1.00	0.97	1.00

Table 1: Correlation coefficients and descriptive statistics for all dependent variables in the study (N = 237). All correlations were significant at p = .001 level. Note: D – DREL test, L or D – its letter or digit version, 2, 3, or 4 – relational complexity level. SPAN – versions of complex span task, L – letter, N – number, S – spatial, F – figural.

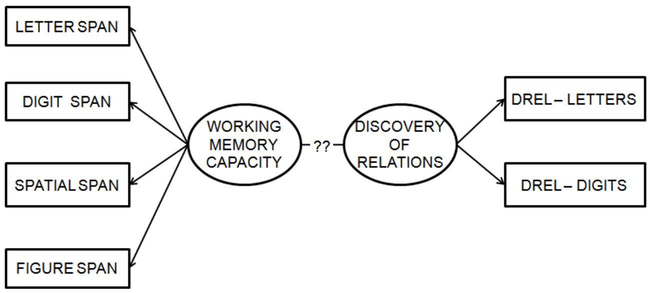

Figure 1: The general structure of the CFA models linking the discovery of binary, ternary, and quaternary relations to WMC. Ovals represent latent variables (factors), while boxes stand for observed variables (measures). Arrows represent factor loadings, while a line stands for correlation.

The two (test versions) by three (levels of complexity) ANOVA of the DREL test's scores indicated that they were significantly higher in the letter version (M = 2.72) than in the digit version (M = 1.77), $F(1, 236) = 316.93$, $p < .001$, $\eta^2 = .57$, and that they decreased with increasing relational complexity ($M_{RC2} = 3.75$, $M_{RC3} = 1.74$, and $M_{RC4} = 1.20$), $F(2, 472) = 1523.80$, $p < .001$, $\eta^2 = .87$. Also, both factors interacted, $F(2, 472) = 43.18$, $p < .001$, $\eta^2 = .15$, as the effect of complexity was more profound in the letter version than in the digit one. These data indicate that the DREL test seems to be a proper tool for measurement of how effectively people discover relations, and that participants were sensitive to the complexity of the test's items.

Then, we tested whether our participants improved at all in the digit version of the DREL test, by comparing their scores on that version to another 79 participants from a similar study, who only attempted the digit version (i.e., they did not "train" on the letter version). This control group scored M = 1.34 per condition (comparing to M = 1.77 in the experimental group), that is, there was a highly significant learning effect, $t(314) = 3.46$, $p < .001$.

Next, with CFA, we assessed the strengths of correlations between the latent variable reflecting WMC (loaded by four complex span tasks) and variables representing the effectiveness of the discovery of relational concepts, separately for each level of complexity. The structure common to three calculated models is shown in Fig. 1. Each model had a good fit, as estimated by Bentler's comparative fit index (CFI; its widely accepted criterion value = .92) and the standardized root mean square residual (SRMSR; the criterion value = .05). For all models, CFIs surpassed .965, and SRMSRs were below .035. Complex span measures' loadings on WMC variable were high (> .667, $p < .001$), as well as loadings of DREL measures (> .609, $p < .001$). This data indicates that the structure of models reflected very well the structure of correlations among variables. The comparison of correlations between both latent variables showed that there was no significant difference between the correlations for binary (r = .663, SE = .068, $p < .001$) and ternary (r = .631, SE = .065, $p < .001$) relations ($\Delta r = -.028$, n.s.), while discovery of quaternary ones was more weakly correlated with WMC (r = .477, SE = .071, $p < .001$) than discovery of both binary ($\Delta r = -.186.$, $t[235] = 2.70$; $p = .004$) and ternary relations ($\Delta r = -.154.$, $t[235] = 2.30$; $p = .009$).

Finally, we tested another CFA model, which related the WMC variable to the index of learning that occurred from the letter to the digit version of the DREL test. Because the scores in quaternary conditions approached floor, and thus the difference between them might have poor psychometric parameters, we decided to aggregate indices of learning of ternary and quaternary relations. The model, presented in Fig. 2, had a very good fit (CFI = .979, SRMSR = .035). Most importantly, it suggests that the performance of participants displaying more capacious WM deteriorated less on the more abstract version of the test in comparison to less capacious participants (r = .207, p = .002), most probably due to a more effective process of the transfer of the abstract pattern of relations, which had been introduced in the letter part of the test, to its digit version.

Discussion

The newly designed DREL test appeared to be a very reliable tool, and scores on DREL responded well to experimental manipulations. The significant drop of the DREL-WMC correlation only for quaternary relations (in comparison to binary and ternary ones) seems to provide more support for Doumas et al.'s (2008) model than to Halford et al.'s (1998) model. Moreover, not only quaternary relations were very difficult to learn (24.5% accuracy), as the latter model predicts, but also ternary relations were rarely found (34.8%), though according to that model they should well fit in WMC of most of participants. In contrast, people displayed fair performance only in cases of binary relations (75.0%), and that fact better corresponds to Doumas et al.'s (2008) assumption telling that during relational learning (but not when processing relations) even as few as two role-filler representations may occupy the whole available capacity. The study provided data convergent with Lewandowsky (2011) results, though moving beyond ternary relations to newly introduced quaternary condition suggests that relational learning is not uniformly linked to WMC with regard to the complexity of relations being learned.

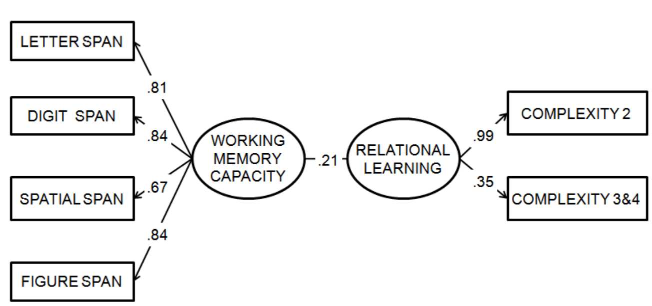

Figure 2: The CFA model linking WMC to relational learning (a difference in scores between two parts of DREL). The same graphical symbols were used as in Fig. 1.

It must be acknowledged, however, that due to the floor effect in the quaternary condition, a possible alternative explanation of the drop in the value of the DREL-WMC correlation coefficients might appeal to a possible worse psychometric usefulness of scores on quaternary relations. However, this is an unlikely explanation, because of relatively low values of the 95% confidence intervals [.338 – .616] for that correlation, comparable to the respective intervals regarding binary and ternary conditions, indicating that all three correlation coefficients have been estimated with a similar precision. Nevertheless, in order to be able to draw firm conclusions on the issue of which model best explains WM contribution to relational learning, the present results should be replicated with a similar method, but one yielding relatively higher scores in the quaternary condition.

Another new result brought by the present study pertains to the fact that not only some general ability to discover relational concepts correlated – though with a varied strength depending on the complexity of those concepts – with WMC, but WMC predicted also the amount of transfer of relational knowledge from one task to another. Although whole our test was strongly dependent on WM resources, we accounted for this fact by subtracting the initial (i.e., general) performance on the task, from the final performance, thus measuring the sheer increase in effectiveness of relational thinking during the coping with the test. It appeared that more capacious WM allows for better learning of abstract relational structures and more effective application of them to new, but analogous, situations. This observation seems to be an interesting challenge for existing models of analogy-making and relational learning, and has potentially profound practical (e.g., educational) implications.

Summary

This study provided another evidence for the thesis that mechanisms of WM impose substantial constraints on human complex cognition, especially its core component: relational thinking. Understanding those constraints by developing computational models of *thinking within WM* is one of the crucial current focuses in cognitive science. This study seems to contribute to those efforts by presenting data supporting those models (e.g., Doumas et al., 2008) which predict that WM resources may be exceptionally loaded during the acquisition of relations, in comparison to a lesser load predicted in situations requiring only transformations and manipulations of relational representations which have already been learned.

Acknowledgments

This work was sponsored by The National Science Centre (NCN) of Poland (grant no. N106 417140). We thank Krzysztof Cipora, Dominika Czajak, and Jolanta Wojcik for conducting the experiment and gathering the data.

References

Chuderski, A., Andrelczyk, K., & Smolen, T. (2013). An oscillatory model of individual differences in working memory capacity and relational integration. *Cognitive Systems Research*. doi:10.1016/j.cogsys.2012.12.005

Conway, A. R. A., 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,* 769–786.

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

Doumas. L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1-43.

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

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology, 50,* 339-368.

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the lab* (pp. 151-175). Washington, DC: APA.

Goodwin, G. P., & Johnson-Laird, P. N (2011). Mental models of Boolean concepts. *Cognitive Psychology, 63,* 34-59.

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, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803-864.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in Cognitive Sciences, 14,* 497-505.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220-264.

Kemp, C., Goodman, N. D. & Tenenbaum, J. B. (2008). Learning and using relational theories. *Advances in Neural Information Processing Systems, 20.*

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 720-738

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

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, Mª A., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence, 39*, 473-480.

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. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49-75). Oxford: Oxford University Press.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General & Applied, 75,* 1–42.