Neurocognitive Mechanisms of Reasoning Ability: An Oscillatory Model of Individual Differences in Relational Integration*

Tomasz Smolen[†]
Pedagogical University of Krakow

Krzysztof Andrelczyk
Institute of Psychology, Jagiellonian University

Adam Chuderski
Institute of Philosophy, Jagiellonian University

Abstract. Using a biologically plausible computational model of working memory, we investigated the neurocognitive mechanisms of finding a simple relation, namely whether syllables located in one row, column, or diagonal line of the matrix of stimuli end with the identical vowel. Though simple, such a task is however a very important tool for understanding of how people cope with various reasoning tasks, because it predicts individual differences in human reasoning in an excellent way. The common mechanism for both finding relations and reasoning may consist of integration of relations due to processing of multiple asynchronous oscillations. It was shown that the distribution of individual ability to integrate relations depends on the level of control over model's internal inhibition, which determines the number of oscillations the model uses to encode complex relational structures.

1 Introduction

Although logicians have categorized reasoning onto several well-defined types, including deduction, induction, and abduction, recent cognitively-oriented theories of reasoning assume that all these types are founded on a common process/mechanism (e.g., Johnson-Laird, 2006; Halford, Wilson, & Phillips, 1998). This assumption is supported by studies which

---

[†] Corresponding author, e-mail address: tsmolen@up.krakow.pl

[*] This work was sponsored by The National Science Centre (NCN) of Poland (grant N106 417140).

Logic & Cognition, 2012: 49–58

show that effectiveness of coping with deductive tasks is perfectly correlated with ability to solve inductive problems (e.g., Wilhelm, 2000), as well as it is suggested by studies indicating strong correlations between scores on diverse reasoning tests and estimates of working memory (WM) capacity (e.g., Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008).

WM is a mechanism responsible for active maintenance of crucial information required for a current task/goal. Most probably, its capacity is related to the number of representations that can be simultaneously held in its most accessible part called primary memory (PM). The most surprising finding is the fact that PM is able to store in parallel only four items on average, and its capacity seems to vary in population from two up to six items (Cowan, 2001). Many interesting formal models of reasoning processes and their underlying WM mechanisms have been proposed in literature (e.g., Hummel & Holyoak, 2003; Carpenter, Just, & Shell, 1990). Unfortunately, formal modeling of such phenomena is very difficult because of their complexity and the related presence of task-specific noise (e.g., differences in strategies and motivation, etc.). So, existing models hardly allow for identification of the specific constraints of PM, which may determine the effectiveness of reasoning.

However, Oberauer, Süß, Wilhelm, and Wittmann (2008) have recently shown that simple tasks which require detection of certain relations among perceptually available stimuli (e.g., testing if the three rhyming words appear in a row, column, or diagonal line of the three-by-three matrix of words) predict both WM capacity and scores on reasoning tests in an excellent way. Oberauer et al. (2008) suggested that the crucial PM constraint on both finding relations and reasoning consists of so called relational integration. PM capacity seems to limit the number of temporary bindings that can be flexibly set between objects and their corresponding roles within some mental structure. Due to temporary bindings, completely new relational structures can be integrated, which seem to be crucial for both deductive reasoning (e.g., integration of the premises into mental models, Johnson-Laird, 2006) and inductive thinking (e.g., mapping of the elements of source and target during analogy-making, Hummel & Holyoak, 2003). But how such an integration is being carried out?

Several models of PM-based integration (e.g., Hummel & Holyoak, 2003; Raffone & Wolters, 2001) describe this process as using a specific pattern of oscillations which encodes a relation being integrated in PM. Specifically, objects and role in a relational structure are bound together by oscillating in synchrony, while distinctive bindings are separated by asynchronous oscillations. This approach in a natural way explains the severe capacity limits regarding relational integration: as brain uses temporal coding for separating representations, and time is a very limited resource, brain is not able to pack too many oscillations into one interval, because they will start to overlap and so the integrated relation will collapse (i.e., the roles of objects in a relation can no longer be identified).

This paper presents a new oscillatory model of PM intended to identify the constraints of reasoning ability. So far, we found out that the model's capacity to keep a large number of items in PM depends on such a level of lateral inhibition that allows relatively many asynchronous but still distinctive oscillations to be maintained. Manipulation in the amount of inhibition allowed us (Andrelczyk, Smolen, & Chuderski, 2012) to account for the distribution of PM capacity in a large sample of participants (see next sections). Because relational integration seems to be a much simpler process then the full-fledged reasoning in complex tasks, while the former is still an excellent predictor of the latter, then modeling the performance on relational integration tasks can contribute to understanding of the nature of constraints pertaining to reasoning.

Thus, the main aim of the present work is to verify if our model is able to directly (i.e., basing on the same values of inhibition parameter which led to the simulation of PM capacity distribution) replicate individual differences in the effectiveness of integration of relations in PM. If the variability in how people can integrate relations and structures can be explained in terms of oscillatory mechanism, it can shed light on probable cause of individual differences observed in reasoning ability, and it may uncover mechanisms responsible for strong correlations between scores on reasoning tests and WM tasks.

2 Oscillatory model of primary memory

The main part of the model is a buffer which can contain a set of processed elements. Every element has three features: (a) reference, (b) output, and (c) level of activation. A reference is a pointer to some content of long term memory, which this element represents (e.g., an object's attribute, a concept, a word, a picture). An output of the element represents its level of its availability to other processes. Finally the level of activation is used for activating or inhibiting other elements in PM.

The output y of the element i in time t has been defined using an unimodal sigmoid function of element's activation x, according to the following formula (1):

$$y_i(t) = \frac{1}{1 + \exp\bigl(-\delta(x_i(t) - \frac{1}{2})\bigr)}. \tag{1}$$

Parameter δ controls the level of nonlinearity of the relation between y and x. If δ is equal to one, then the relation is linear. When δ approaches infinity, then the relation between y and x becomes threshold function switching at $x = \frac{1}{2}$. Moderate δ values give a graded threshold function.

The level of activation x of the element i in time t is computed on the basis of equation (2)

$$\begin{aligned} x_i(t) = \quad & x_i(t-1) + \frac{\lambda}{1 + y_i(t-1)} + \alpha \sum_k \exp\bigl(x_k(t-1) - x_i(t-1)\bigr) \\ & - \beta \sum_j \exp\bigl(x_j(t-1) - x_i(t-1)\bigr) + \varepsilon(n). \end{aligned} \tag{2}$$

Parameter λ controls how much an element i will be auto activated by the recurrent connections feeding its output back into it, what reflects the commonly postulated self-recurrent nature of neuronal groups in brain structures underlying the focus of attention (e.g., O'Reilly & Munakata, 2000). Parameter λ primarily regulates the frequency of oscillations.

Index k denotes elements which output just before element i does, more precisely those which have slightly higher activation (in range $[y_i, y_i + \kappa]$) than i has. Parameter α determines how much the output of elements, which oscillate close to element i, will increase its activation. This accounts for the known fact that neurons which fire in synchrony with a given neuron strongly influence its potential. Such a mechanism of coactivation helps to maintain synchrony among items with similar outputs. Parameter κ defines also the temporal resolution of bindings: the larger is κ value, the more distant (in terms of activation) elements, namely those in $[y_i - \kappa, y_i + \kappa]$ range, will be considered by the model as bound to the same part of relational structure (e.g., to the same premise)

Index j denotes elements which are not k nor i elements, namely those that fall out of the aforementioned range. These elements encode representations which are separate from a representation encoded by elements i and κ. Parameter β controls the strength of inhibition exerted by elements j, namely how much they will decrease the activation of element i. This strength depends on a difference in the activity of elements i and $j$activity: relatively more active element j will inhibit element i more strongly than will do a less active j. The last term of equation (2) consists of noise (ε), which is being drawn from the normal distribution with a mean equaling zero, and a variance dependent on parameter n.

The activations and outputs of elements are updated in discrete cycles. Each cycle represents a period of several milliseconds, though any precise timings of the model's operation are not being reported in this paper. As soon as an output of an element reaches unity (this reflects firing of a neuronal group), the parameter λ for that element is temporarily changed to a relatively large negative value, which makes this element quickly fall below a base level of activation (set to .2). This is meant to reflect the phenomenon of refraction. Then, the value of λ is being reset to a default value and the element starts building up its activation above the base level. However, in some cases, inhibitory signals may be so strong that the activation may reach zero and the element may permanently fall out of the buffer.

In the model, the number of elements that can be bound together within one synchronous oscillation is not limited. However, in the following simulations only pairs of synchronized elements were used. Each pair consists of an item that is maintained in memory, and a context in which it was encoded.

Maintenance of bindings in WM is successful if neither they fall out of the buffer (i.e., their activation stays higher than zero) nor their oscillations group with other bindings . After the set of elements have been processed, it can be either actively removed from memory or left there to fall out when subsequent bindings are added. Bound elements (pairs in case of research described herein) are being added to the buffer in the same time. The first pair is being added at a random level of activation. Subsequent pairs can be added when an activation of all other elements is lower than $1 - 4\kappa$. New elements are being added at a level of activation $x = x_{\max} + \kappa + (1 - x_{\max})/2$, where $x_{\max}$ denotes x value of the most active element. This mechanism grants that at least on entering the buffer new elements are sufficiently distinctive from all other elements.

In the model, the capacity limit arises because adding consecutive pairs increases the strength of total inhibition. When the value of inhibition, which an element receives, surpasses the value of auto-activation (regulated by parameter λ) and co-activation (governed by parameter α), the elements with the lowest activation levels start falling out of the buffer. If one element from the pair falls out, then the coactivation is no longer possible, and the chance that the other element from that pair would also fall out drastically increases. Thus, the parameter β is the main determinant of the model's capacity. The higher value parameter β has the faster the elements start to fall out of the buffer. So, the model predicts that a maximum capacity will be achieved when there is no inhibition at all ($\beta = 0$). Indeed, simulations showed that in such a case the model was able to maintain twelve pairs, surpassing human capacity, but only when the noise was switched off ($n = 0$). In more realistic cases, a certain amount of inhibition is necessary because it secures that oscillations will evenly occupy a respective time interval. Thus the optimal values of β reflect the trade-off between low inhibition (many items can be maintained, but there are no bindings nor selective removal) versus high one (less items can be maintained, but they are properly bound). By gradually increasing moderate value of β, the highest capacity (around five items) observed among people was replicated, as well as mean capacity (about three items), and – the lowest possible capacity (one item). Respective patterns of oscillations are presented in fig. 1.

Although the capacity of the model is primarily determined by parameter β, it can be noted that three other parameters can in principle modulate workings of the model. Firstly, the increase in parameter α makes bounded elements more strongly synchronous. Such a mechanism may reflect a top-down boosting applied by the prefrontal cortex, which can pass additional activation to PM (see Edin et al., 2009). This extra boosting makes elements firing in temporal proximity more active. Another factor which impacts capacity is the level of noise (n). The higher noise, the higher is the probability that pairs get desynchronized. The noise may reflect numerous distinct factors, as fatigue, mental retardation, influence of drugs, etc. The last parameter related to capacity is the value of κ parameter. If κ is large, then the pairs are stable, but there is little "room" for adding new pairs and the capacity is low. If κ is very low, in theory many pairs could be added, but even tiny differences in paired elements' activity make pairs desynchronize.

3 Simulation of individual differences in primary memory capacity

The model is able to recreate numerous effect regarding PM. For brief overview see (Andrelczyk et al., 2012). The crucial effect recreated by the model is the distribution of PM capacity. The estimation of individual PM capacity of 168 participants was based on their performance in a two-array comparison task (Luck & Vogel, 1997, see Andrelczyk et al., 2012, for details).

The original task required memorizing an array of a few items. Then, after a retention interval, the array was repeated, but there was 50% chance that one of the items was changed. The task was to indicate whether the item had changed or not. A version of the task consisting on a single-probed recognition was used: one of the items in the second array was surrounded by a cue indicating that if any of the items had changed, it was the surrounded one. The PM capacity k is estimated on the basis of proportions of correct and

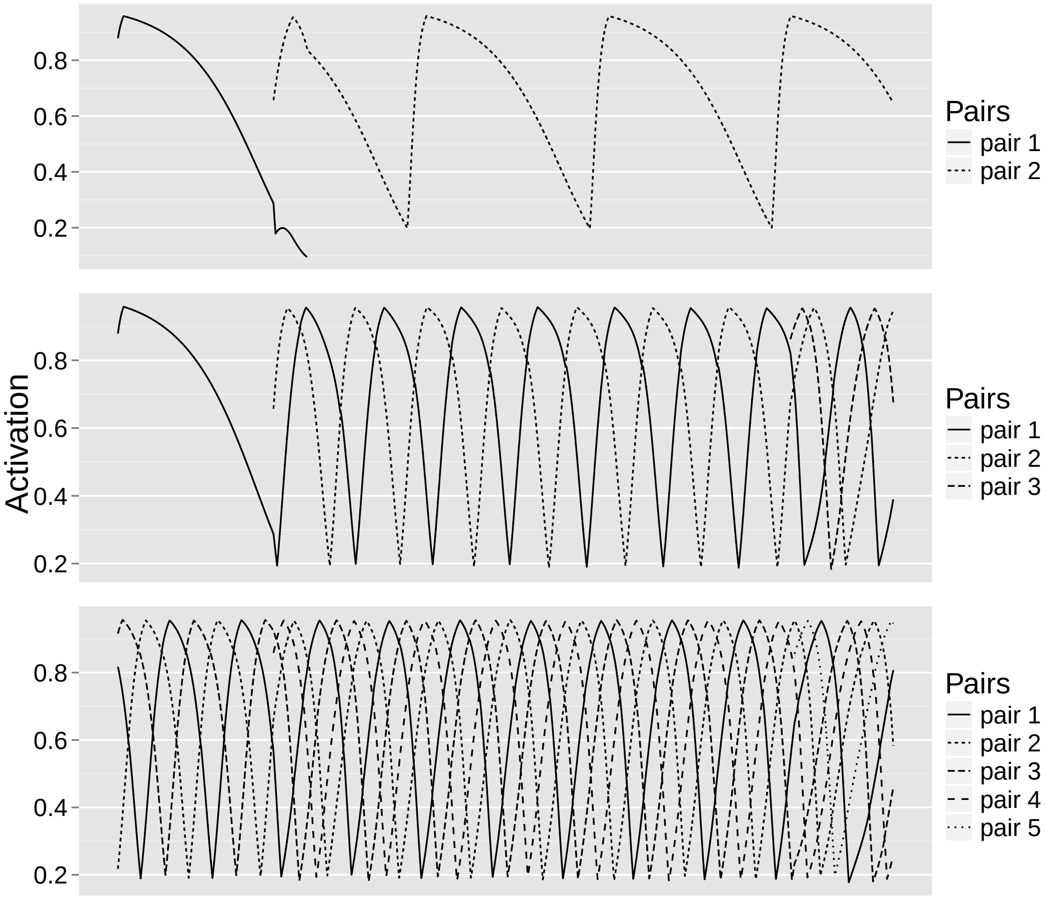

Figure 1. Patterns of oscillations for the lowest (one binding; upper panel), medium (three bindings; middle panel), and largest (five bindings; bottom panel) capacity. When k is insufficient, addition of a new pair eliminates an existing pair (see upper panel). When pairs' oscillations peaks are close they tend to achieve equal intervals (see middle panel).

incorrect responses (both misses and false alarms) in conditions which require different PM capacity to generate correct response (i.e., different array sizes). Simulated results were generated by running 168 versions of the model with individual β values drawn from the normal distribution ($\mu = .026$, $\sigma = .004$), which granted the best fit. Comparison of results of the experiment and simulation are shown on fig. 2.

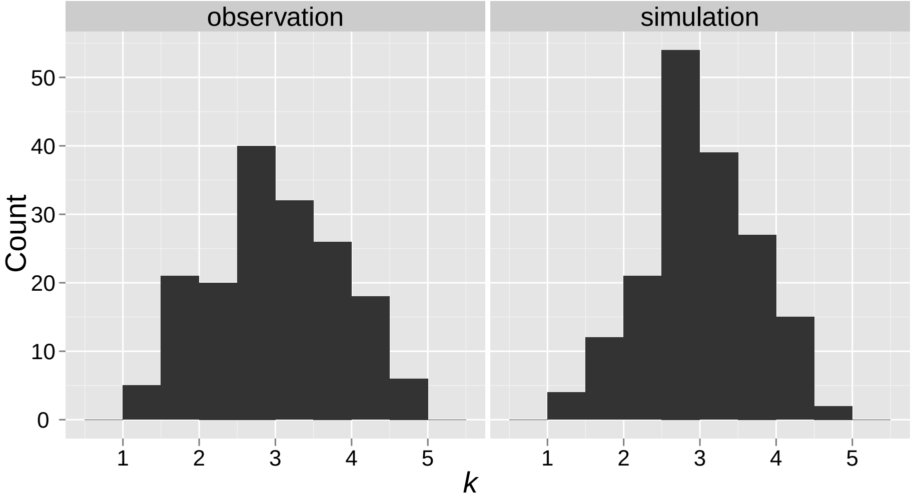

Figure 2. Histograms of k values derived from the two-array comparison task in case of human subjects (left panel) and simulated data (right panel).

4 A basic mechanism of reasoning: relational integration

In order to measure relational integration, the monitoring task introduced by Oberauer et al. (2008) was used. The task consisted of the presentation of 80 patterns. As much as 20 patterns included stimuli that fulfilled some simple relation defined in the task's instruction. The patterns consisted of a three by three array of two-letter syllables. A participant's goal was to detect the specified relation in each array. Two subsequent arrays differed by exactly one syllable; a short "blink" of the array indicated that a syllable had been substituted. Participants were to look for three syllables ending with the same vowel, located in one row, column, or diagonal line. Level of performance was calculated as a number of correct responses for the specified relation minus one third of false alarm errors (as three times more non-target patterns occurred than target patterns).

The model of the relation monitoring task was based on assumptions that people systematically seek for relation and that possible differences in strategies can be neglected with no effect on efficiency of the general performance. These assumptions were partially justified by empirical results (see Chuderski, Taraday, Necka, & Smolen, 2012). The model searches for relations sequentially, first in rows, then in columns and diagonal lines. Every test consisted of two steps: (a) check if two examined elements satisfy the required condition (namely if the last vowels of the syllables match), and if so (b) test if the third element satisfies that condition. If the second stage gives positive result, then the search is stopped and response is generated. If it does not, a next row/column/diagonal line is processed. Such a procedure was repeated for each matrix pattern of stimuli.

The key concept in this article is relational integration. As it was shown above, the difference between maintaining elements in PM as parts of a relation, and just storing them in memory in unrelated way, is crucial for explanation of reasoning ability. In the model described above, several elements are integrated into a relation if they are represented in PM and at the same time can be associated with some roles in that relation. The elements can be interpreted as the arguments of a relation as long as they can be identified as playing a well-defined role in the structure of that relation. This is done by binding them to particular role-defining tags (e.g., predicates), with use of synchronous oscillations. For example, in the model of the monitoring task, a relation of being the identical letters in one row (column, diagonal) is considered. In such a case, each element in PM must be bound to a tag which informs about a location within a matrix at which the element was placed. If tags of the elements held in PM inform that the elements lay in one row (column, diagonal line), then the rule testing if they are identically ended can be fired and a response can be made on a basis of the rule's result.

5 Zero-parameter cross-task simulation of individual differences in relational integration

Zero-parameter cross-task modeling is a technique of testing model's validity. A model is considered to be justifiable if it achieves good fit to data in new task, given parameters' values matched in some another task. The main goal of this article is to show that the model is able to recreate results in the monitoring task when provided with exactly the same values of β parameter, as estimated in the two-array comparison task.

In the model of the monitoring task, we also used the same values of all other parameters as in a previous model. The parameter α was set to a value of .0001, κ – to .05, and n – to zero, (for rationale see Andrelczyk et al., 2012). Values of λ were drawn from a normal distribution which optimised the model's capacity given the adopted range of parameter β.

Three properties of resulting distributions of accuracy in the monitoring task were compared: mean, variance, and skewness. The mean proportion of correct responses in observed results was .8, 95% bootstrapped confidence intervals (BCI, 10000 samples) [.78, .82], and in simulated results – .8, 95% BCI [.79, .82]. The variance equaled .013 (95% BCI [.01, .015]) and .012 (95% BCI [.01, .014]), respectively. Finally, the skewness of observed results distribution was $-.3$, 95% BCI $[-.5, -.09]$, while for the simulated results it was equal to $-.37$, 95% BCI $[-.57, -.16]$ (see fig. 3.). The difference between these two distributions was not statistically significant ($\chi^2 = 9.02$, $df = 9$, $p = .43$) The simulated results predicted observed data well ($r = .85$, 95% CI [.47, .96]).

Correlation between the estimated value of k (PM capacity) and correctness in monitoring task was moderate ($r = .36$, 95% CI [.22, .48]) and very close to the observed one ($r = .33$). Such correlation value is consistent with suppositions considering relations between two examined abilities – relational integration ability and PM capacity. It was assumed that relational integration highly relies on PM capacity but they are not the very same construct.

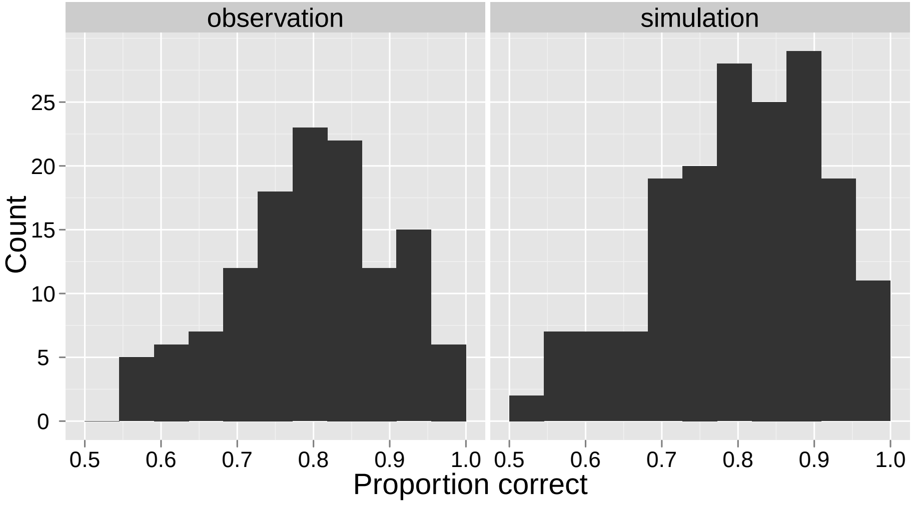

Figure 3. Histograms of proportion correct in observed (left panel) and simulated (right panel) results regarding the relation monitoring task.

It must be took into account that the above said slight correlation is partially caused by limited reliability of both tasks.

6 Discussion

To sum up the results, we applied an exact variation in the parameter reflecting internal inhibition, which previously allowed us to computationally replicate the indicators of PM storage capacity observed in the human subjects, to the relation monitoring task, and in result we recreated the individual differences in the ability to find relations in patterns of stimuli. As the relation monitoring task is an excellent predictor of reasoning ability in general, factors limiting the performance in that task may be valid candidates for causes of substantial differences in coping with reasoning problems, as observed in human population.

One such plausible factor is the level of control over lateral inhibition within WM, as demonstrated by our simulation. Human brains which are able to keep the inhibition level low, thus allowing for occurrence of asynchronous oscillations supporting as many distinct elements of a relational structure as possible, may be also able to cope with more complex problems than do brains that control their inhibition level poorly.

The presented work is a very preliminary one so far. The next step that should be made would be building the models that can solve complex reasoning problems (not only finding simple relations) and demonstrating that inhibition level determines how well such models can reason.

References

Andrelczyk, K., Smolen, T., & Chuderski, A. (2012). Oscillatory basis of individual differences in working memory capacity. In N. Rußwinkel, U. Drewitz, & H. van Rijn (Eds.), *Proceedings of 11th International Conference on Cognitive Modeling* (pp. 181–186). Berlin: Technische Universität Berlin.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the raven progressive matrices test. *Psychological Review*, *97*, 404–431.

Chuderski, A., Taraday, M., Necka, E., & Smolen, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295.

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs but why? *Intelligence*, *36*, 584–606.

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

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegner, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *PNAS*, *106*, 6802–6807.

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.

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

Johnson-Laird, P. N. (2006). *How we reason?* Oxford: Oxford University Press.

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

Oberauer, K., Süß, H. H., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641–642.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience.* Cambridge MA: MIT Press.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual memory. *Journal of Cognitive Neuroscience*, *12*, 766–785.

Wilhelm, O. (2000). *Psychologie des schlussfolgernden denkens: Differentialpsychologische prüfung von strukturüberlegungen.* Hamburg: Dr Kovac.