Oscillatory Basis of Individual Differences in Working Memory Capacity

Krzysztof Andrelczyk (kandrelczyk@gmail.com)
Institute of Psychology, Jagiellonian University
Mickiewicza 3, 31-120 Krakow, Poland

Tomasz Smolen (tsmolen@up.krakow.pl)
Pedagogical University
Podchorazych 2, 30-084 Krakow, Poland

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

Abstract

The paper presents a novel formal model of the active buffer of working memory. The model uses synchronic oscillations in order to bind an item and its corresponding context into one representation, while asynchronic oscillations allow the model to maintain several separate representations. Due to bindings, the model exerts proper control over the buffer's contents, as demonstrated by effective rejection of distractors. Most importantly, the model predicts an inherent limitation in WM capacity that arises from the trade-off between the number versus the stability of representations bound by oscillations, which depends on the strength of lateral inhibition present among oscillating items. The systematic variation in inhibition leads to exact replication of the capacity distribution observed in a large sample of participants, as well as to prediction of a few novel, capacity-related experimental effects.

Introduction

Working memory (WM) is a neurocognitive mechanism responsible for the active maintenance of information for the purpose of its ongoing processing. It plays a crucial role in many complex cognitive processes like relational reasoning, problem solving, language, and learning (Jarrold & Towse, 2006). One of the most important features of WM is its heavily limited capacity. Usually, a person can maintain two up to six items in WM, with a mean individual capacity equaling four items (Cowan, 2001).

Some theories (Cowan, 2001; Oberauer, Süß, Wilhelm, & Sander, 2007) predict that WM consists of two distinct structures: a highly active and accessible buffer called *the focus of attention* (or *primary memory*, PM), and a less accessible *activated long-term memory* (or *secondary memory*, SM). It is argued that only PM is capacity limited, while SM is not, and that this very limit influences human performance on various tasks.

The most promising theoretical approach to storage in PM explains it as some kind of a pattern of oscillations. Several oscillatory models describe in a neurally plausible way the PM mechanisms which use patterns of fast, repetitive changes in activity of stored representations (i.e., use more than one oscillation during retention time) for coding items (Edin et al., 2009; Horn & Usher, 1992; Jensen & Lisman, 1998; Usher, Cohen, Haarmann, & Horn, 2001) and binding together different features of a maintained item (Hummel & Holyoak, 2003; Raffone & Wolters, 1998). Such models generated numerous predictions supported by neuroimaging data (e.g., Edin et al., 2009; Jensen & Lisman, 1998; Raffone & Wolters, 1998). They also showed that temporary bindings are crucial for complex cognition because they allow for representing arbitrary relational structures (Hummel & Holyoak, 2003; Oberauer et al., 2007).

Finally, oscillatory models (e.g., Hummel & Holyoak, 2003; Jensen & Lisman, 1998; Raffone & Wolters, 1998; Usher et al., 2001) naturally explain capacity limits as an emergent property of PM, which results from the trade-off between the number of to-be-maintained representations versus the ability to distinguish among them. As brain uses temporal coding for separating representations in PM, and time is a very limited resource, brain is not able to pack too many oscillations into one interval, because they start to overlap and so they stop being distinctive (=informative).

On the contrary, in models which do not rely on oscillations, one has to set a limit on the number of PM's slots (e.g., Kahana, 1996) or the amount of PM's activation (e.g., Daily, Lovett, & Reder, 2001) in an arbitrary way, so no natural capacity limit is being explained. Similarly, models which use only one cycle of activation change to code an item (e.g., Botvinick & Plaut, 2006; Davelaar, Goshen-Gottstein, Ashkenazi, Haarman, & Usher, 2005), seem to be less neurally and functionally plausible than the full-fledged oscillatory models.

Although the oscillatory models nicely explain how brains handle maintenance in PM, and they give important insights into the nature of capacity limits, no such model has yet dealt with the fact that *people differ in capacity*. Though in principle we all could have had the maximum possible capacity, in fact capacity is hugely varied among humans.

This paper presents a novel formal model of PM. It aims to demonstrate which features (i.e., parameters) of the model's oscillatory mechanism are responsible for the observed individual differences in PM capacity. We test if a systematic manipulation to one such parameter, namely the strength of lateral inhibition applied among memory items, can replicate the distribution of capacity estimates in human population as well as a number of more specific effects.

Oscillatory model of primary memory

The main part of the model is a buffer, which contains a certain number of elements. Each element roughly approximates a neuronal assembly representing one specific feature of the world (e.g., an object's attribute, a concept, a word). As in many other models, a level of internal activation x_i, which falls in [0, 1] range, is assigned to each element i.

The external output y of the element i in time t has been defined using a commonly applied sigmoid function of x_i, according to the following formula (1):

$$y_i[t] = \frac{1}{1+\exp(-\delta(x_i[t]-.5))}$$

Parameter δ controls the level of nonlinearity of the relation between y and x. For small δ values this relation for $0 < x < 1$ is almost linear. With increasing δ, (1) gradually alters into a threshold function with the threshold at $x = .5$.

In order to express the presumed mechanism responsible for binding the features of one item while keeping the different items disjoint, we introduced a completely new (i.e., in comparison to other existing oscillatory models) equation, which controls changes in levels of activation (2):

$$x_i[t+1] = x_i[t] + \frac{\lambda}{1+y_i[t]} + \alpha\sum_k \exp(x_k[t]-x_i[t]) - \beta\sum_j \exp(x_j[t]-x_i[t]) + \varepsilon(n)$$

Parameter λ controls how much element i is autoactivated by the recurrent connections feeding its output back into it, what reflects a 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, namely those in $[y_i, y_i + \kappa]$ range. So, parameter α determines how much the outputs of elements, which oscillate close to element i, 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 κ, the more distant (in terms of activation) elements will be considered by the model as bound within the same representation, namely those in $[y_i - \kappa, y_i + \kappa]$ range.

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

The activations and external outputs of elements are updated in discrete cycles. Each cycle represents a period of several milliseconds, though precise timings of the model's operation were not reported in this paper. As soon as 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 in the model 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, inhibition signals may be so strong that the activation may decrease below zero – a value adopted as a minimal activation necessary to stay in the buffer – and the element permanently falls out of it.

Generally, the number of elements that can be bound together within one synchronic oscillation is not limited. However, in the following simulations we apply only pairs of synchronized elements (an item identity and its position).

Workings of the oscillatory model

The aim of the model is to maintain as many separate oscillations as necessary, for a given interval. Two elements making one oscillating pair (e.g., a letter and its temporal or spatial position, see below) are added to the buffer in the same time. The first pair is added with a random level of activation. Subsequent pairs can be added when activations of all other pairs $\forall x < 1 - 4\times\kappa$. They are being added at a level of $x = x_{max} + \kappa + (1 - x_{max}) / 2$, where x_{max} denotes x value of the most active pair. This mechanism checks if there is enough place in activation space for new elements, and grants that at least on entering the buffer new pairs will be sufficiently distinctive from all other pairs.

In the model, the capacity limit arises because addition of consecutive pairs increases the strength of total inhibition that each pair receives. When this value surpasses the results of autoactivation (regulated by parameter λ) and coactivation (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 β, the faster the elements start falling out of the buffer. So, the model predicts that a maximum capacity will be achieved when there is no inhibition at all (β = 0). Indeed, in such a case, the model was able to maintain twelve pairs, surpassing human capacity, but only when the noise was switched off. In more realistic cases, a certain amount of inhibition is necessary because it secures that oscillations will evenly occupy a respective time interval, helping to separate them. So, the most appropriate values of β reflect the trade-off between low (many elements can be maintained, but they are unbound) and high (less elements can be maintained, but they are properly bound) inhibition.

By gradually increasing the moderate value of β, we replicated the highest capacity (around five items) observed among people (see the next section), mean capacity (around three items), and – the lowest possible capacity (one item). Respective patterns of oscillations are presented in Fig. 1.

Although β is the most important determinant of the model's capacity, we note that three other parameters can in principle modulate workings of the model. Firstly, the increase in parameter α would strengthen synchrony of bound elements. 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 all elements more strongly activate each other. However, because we assume that a given element, when fires, can activate only elements firing in its temporal proximity, we expect that the boosting influences only the mechanism of coactivation. 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. 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 the pair desynchronize.

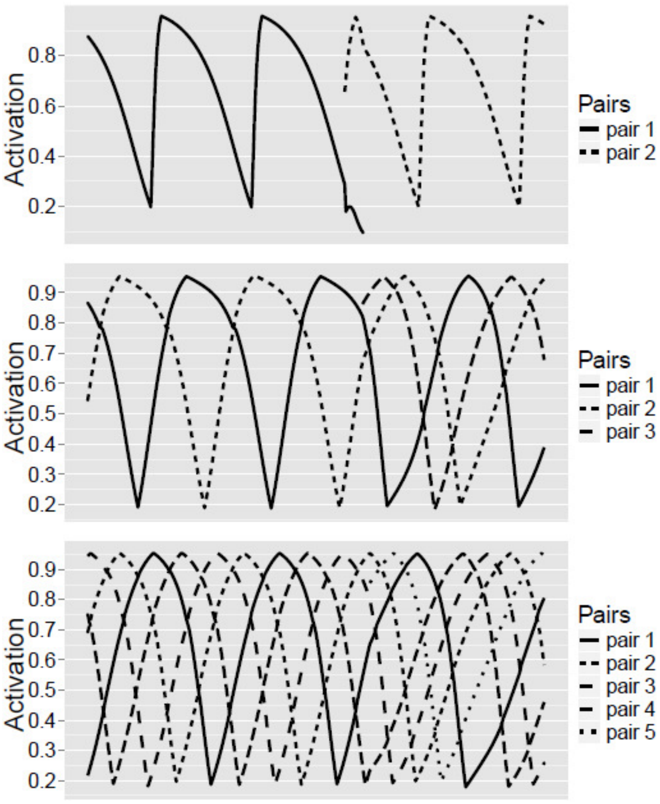

Fig.1: Patterns of oscillations for the lowest (=one item; upper panel), medium (=three items; middle panel), and largest (=five items; bottom panel) capacity. When capacity was insufficient, addition of a new pair eliminated an existing pair (see upper panel).

Measurement and simulation of individual differences in primary memory capacity

In the following simulations, we set α parameter to a low arbitrary value of .0001. Regarding κ, changing its values between .03 and .07 did not influence the model's capacity, so we set the κ value to .05. Parameter *n* was set to zero (i.e., noise was turned off). Parameter λ was drawn from a normal distribution which optimized the model's capacity given the adopted range of parameter β. So, in total we adjusted four global parameters. In order to replicate the distribution of WM capacity as observed in a sample of participants, we individually varied the values of β (see below).

We modeled two similar WM tasks. In the first one (the Sternberg task), the model attempted to add to its buffer several letter-position pairs, and then a probe in a particular position was presented. The model ran two processes which tested (a) whether the element identical to the probe could be found in the buffer and if a position bound to it (if any) matched the position of the probe or not, and (b) whether the element identical to the probe's position could be found and if an element bound to it matched the probe's identity or not. If either the identity or the position was found, and its binding matched either the probe's position or identity, respectively, then the model generated a positive answer. If both elements were found, but one of them did not match the corresponding element, the answer was negative. If any of two elements were not found, the model guessed either the positive answer with probability ρ (a decisional bias) or a negative answer with a chance 1 – ρ.

Basic effects regarding primary memory

We started testing our model by checking if it is able to replicate the recency effect. When WM performance relies primarily on PM, as for example in the Sternberg task applied with fast presentation rate, an increased accuracy is observed for the most recent items in comparison to middle items, but there is no primacy effect (Chuderski, Stettner, Orzechowski, 2007). The simulated effect is shown in Fig.2.

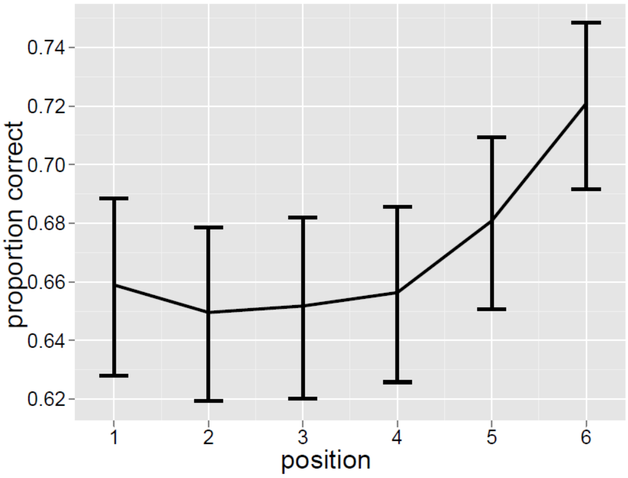

Figure 2: Accuracy of the model in all serial conditions of the six-item Sternberg task.

A more interesting observation – in light of the aim of this paper – regards the fact that people quite effectively use positional information to reject distractors presented to them. For example, one of us (Chuderski & Stettner, in revision, Exp. 1, positive digit condition) used a modified Sternberg task, which was analogical to the standard version with one exception that a probe (a letter) was accompanied by a digit, which denoted the letter's position in a memory set. The digit could match the target's position or not, and the task of 47 participants and 47 corresponding simulations was to accept only matching digits. We observed that participants correctly accepted more matching digits (M = .78, SD = .11) than incorrectly accepted non-matching ones (M = .26, SD = .17). This result indicates that they effectively maintained the positions of items in WM. Simulated results were close to observations: model accepted M = .67 of the matching digits, while it did not reject only M = .33 of the to-be-rejected digits. Slightly lower accuracy of the model resulted from the fact that it only used its PM, while people most probably relied their performance on both PM and SM. This result indicates that a proper PM model must account for binding of the representations with their contexts. An unbounded information may often be simply useless.

The distribution of primary memory capacity

The crucial simulation consisted of the replication of the distribution of PM capacity estimates, which had been observed in the sample of 168 young participants, who fulfilled a two-array comparison task (Luck & Vogel, 1997). The task is assumed to require maintenance of material in PM, while SM barely helps in doing this task due to the use of figural material and a fast presentation rate. Ten other participants were excluded from the original data because their results in the task suggested that they did not succeed to maintain even one item in their PM.

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 item was changed. The task was to indicate if the item had changed or not. We used a version of the task consisting on a single-probed recognition: one of the items in the second array was surrounded by a cue indicating that if any of the items had changed, it was only the surrounded one. The test included 90 trials. Each self-paced trial consisted of a virtual, four by four array filled on random with four, five, or six (i.e., set size) stimuli, being drawn from a pool of 16 simple black figures (e.g., a square, a circle, a rhombus, an arrow, a cross etc.), each approximately 2.5 × 2.5 cm in size. The array was presented for set size multiplied by 0.5 s.

An estimate of PM capacity uses the proportion of correct responses for arrays with one item changed (hits; H) and the proportion of incorrect responses for unchanged arrays (false alarms; FA). PM capacity is estimated to k items (out of N items of the set size), on the assumption that a participant produces a correct hit or avoids a false alarm only if a cued item is transferred to his or her PM (with the k/N chance). If a non-transferred item is cued, then a participant guesses the answer. Thus, the sheer PM capacity is equal to $k = N \times (H - FA)$. The value of k is believed to closely approximate the actual number of items held in PM by an individual (Rouder, Morey, Morey, Cowan, 2011).

We used an analogous models as for the Sternberg task, with an exception that this time it encoded figures and their spatial positions. The value of β = .0026 allowed us to replicate the mean k value in the sample ($M\ k_{\text{sim}}$ = 3.01, $M\ k_{\text{obs}}$ = 2.92). In order to simulate 168 individual results we varied values of parameter β for each individual simulation, drawing it from the normal distribution with M = .0026 and SD = .0004. Histograms of the observed and simulated distributions of k values are presented in Fig. 3. Both distributions did not differ significantly (χ^2 = 6.97, df = 7, p = .431). R^2 value for observed and simulated data was .93.

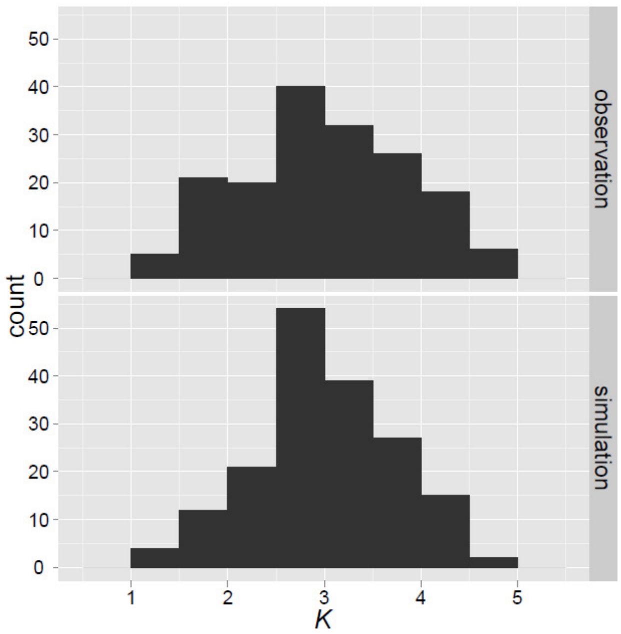

Figure 3: A number (count) of observations (upper panel) and simulations (lower panel) yielding particular k values.

Experimental effects related to capacity

Next, we examined if there were any specific differences in performing the task related to differences in participants' capacity and – if yes – whether the model was able to predict them. Analysis of observed data indicated that participants more accurately responded to unchanged (congruent) arrays than to changed (incongruent) ones. In the model, this was accounted for by setting ρ value to .43. As parameter ρ regards guesses, and as highly capacious participants rarely guess (all necessary information is in their PM), the difference between accuracy in both conditions should diminish with increasing k value. This effect was found in both observed and simulated data (see Fig. 4).

Consequently, the model predicted that the discrepancy in accuracy between the incongruent and congruent conditions would be increasing as N increases (here, from four to six items). Such a pattern have also been found in observed data (see Fig. 5). However, it appeared that the model overpredicted accuracy in the four-items and five-items conditions, while it underpredicted accuracy in the six-item condition.

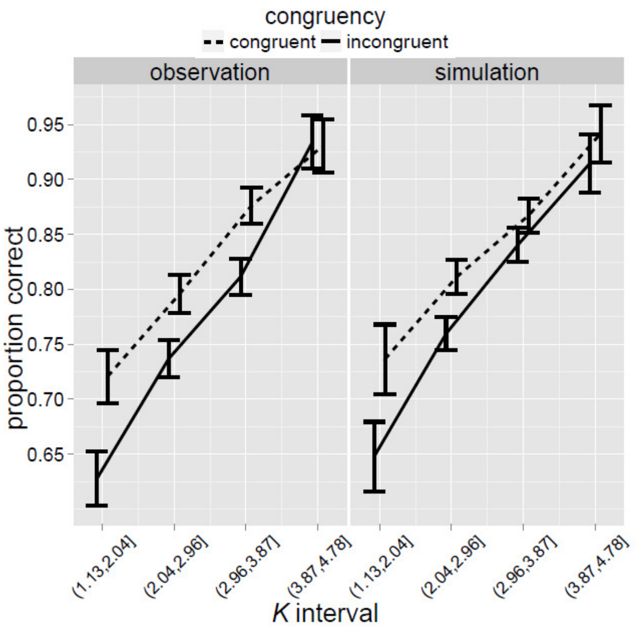

Figure 4: Accuracy in the congruent and incongruent conditions, in a function of the *k* value.

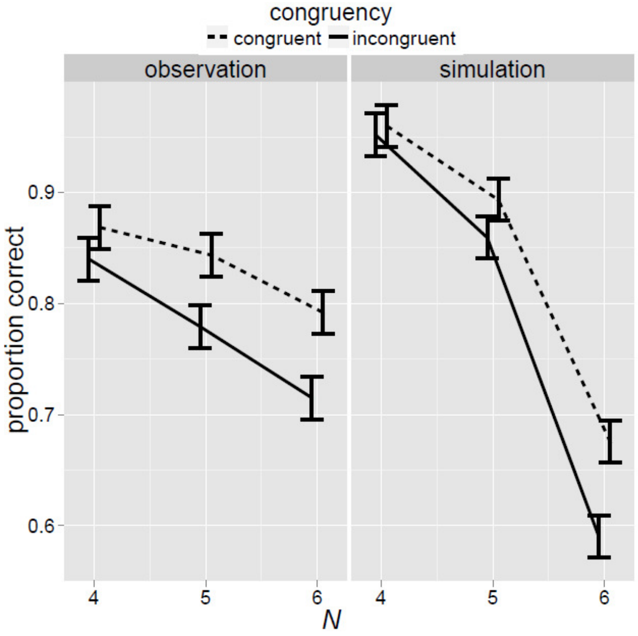

Figure 5: Accuracy in the congruent and incongruent conditions, in a function of the set size (*N*).

In search for a possible cause of the mismatch, we investigated if it can be related to the differences in capacity. We compared the *k* values for the respective item conditions between participants and between simulations yielding high versus low *k* estimates ($k > 3.5$ vs. $k < 2.5$). We found that in case of highly capacious participants, the *k* estimate significantly increased with *N*, $\Delta k = 0.85$, 95%CI = [0.58, 1.12], while in case of the low-capacity group the *k* value was significantly lower in the five-item condition than in the six-item one, $\Delta k = -0.37$, 95%CI = [-0.09, -0.64] (see Fig. 6, left panel). The former effect is a direct consequence of the fact that *k* estimates of some participants equaled or surpassed four, so *k* value in the four-item condition underestimated their capacity. In fact, when only the six-items condition was considered, the maximum human capacity was $k = 5.6$, and was accounted for by the model. The latter effect is much more interesting: it indicates that PM of low capacity persons was even less effective than usual if the discrepancy between their actual capacity and the imposed requirements increased. This result is coherent with other behavioral and neuroimaging data (e.g., Todd & Marois, 2004).

The analysis of how the model coped with increasing *N* depending on adopted value of β determining its capacity, indicates that it showed qualitatively similar pattern of data (see Fig. 6, right panel, black lines), though there were substantial quantitative differences in comparison to observations. Increasing *N* value was on average not disruptive for the model's capacity in case of low values of lateral inhibition, why it dramatically decreased its capacity when the level of inhibition was high (i.e., when *k* was low).

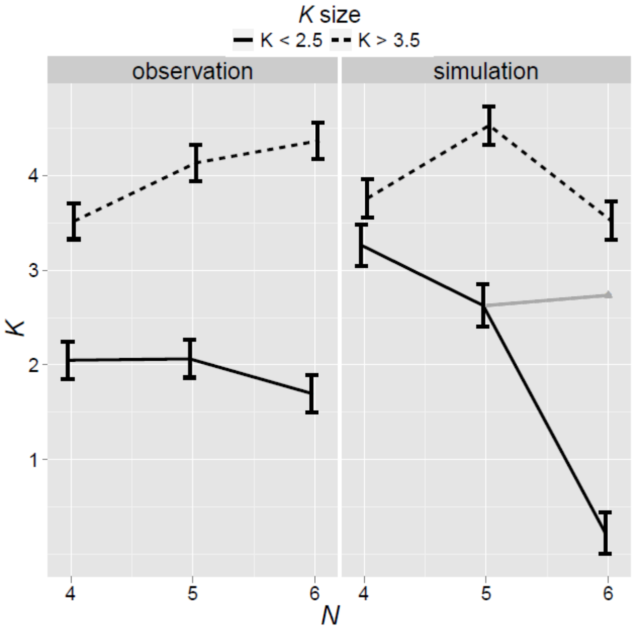

Figure 6: Mean values of *k* in a function of set size, for low-capacity (total $k < 2.5$) and high-capacity (total $k > 3.5$) participants/simulations.

What could be responsible for the model's more profound effect of the discrepancy between applied *N* value and individual *k* value? Edin et al. (2009) suggested that when such a discrepancy occurs, an additional top-down activation is recruited by the brain in order to counteract the lateral inhibition surpassing the brain's capability of dealing with it. We tested this hypothesis by re-running the six-item condition with twice as large α value (.0002) as in the original simulation. In result, the model's accuracy highly increased and our data better fitted human data (see Fig. 6, right panel, gray point). So, most probably our initial setting of α value underestimated the role of autoactivation in PM.

Discussion

Using a novel oscillatory model, which was aimed to reflect the mechanisms of active (attentional) maintenance of information in PM, the presented study has shown that variation in the strength of lateral inhibition among oscillating representations, which is necessary for formation and temporal separation of bindings among these representations, allowed for accounting for individual differences in PM capacity in large sample of participants. In our sample, it varied from around one item up to almost five items. It could even reach almost six items, when the largest set size was considered. This result seems to pose a serious problem to those oscillatory models, which predict that the maximum WM capacity is only four items (e.g., Raffone & Wolters, 1998). Most probably, these models would not be able to mimic the full distribution of WM capacity in human population. On the contrary, due to ability to coactivate the elements oscillating together, the presented model was able to maintain a dozen separate items at maximum, though when its parameters were being set on neurobiologically plausible levels (e.g., there was non-zero lateral inhibition), the model's capacity was naturally constrained to several items. In our view, the study suggests that the brain's ability to control (decrease) the level of inhibition within PM underpins it mechanisms supporting active maintenance of as much separate representations as possible.

However, the story regarding the replication of individual differences in capacity is not that simple. A more precise analysis of the effects of memory load in a function of individual capacity showed that the strength of lateral inhibition is not the only factor influencing the model's capacity. When the model attempted to maintain too much items in relation to its actual capacity, this increased the inhibition to such a high level that it led to a catastrophic decrease in capacity. In line with others (e.g., Edin et al., 2009; Todd & Marois, 2004), we suppose that the main role of the prefrontal cortex in active maintenance of information is to prevent such situations by additionally activating PM in a top-down manner. In our model, this was done by adjusting the coactivation of elements oscillating together. The analysis of computational properties of the coactivation and search for α values enhancing the fits of the model should be the subject of our future investigations.

Summing up, we presented a preliminary but highly original study on the neurocognitive mechanisms underlying the individual variation in PM capacity. Its results suggest that the concepts of oscillations and bindings can have a great explanatory power in regard to working memory.

Acknowledgments

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

References

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order. *Psychological Review, 113,* 201-233.

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

Chuderski, A., & Stettner, Z. Access to working memory and its relation to reasoning. *Manuscript in revision.*

Chuderski, A., Stettner, Z., & Orzechowski, J. (2007). Computational modeling of individual differences in short term memory search. *Cognitive Systems Research, 8,* 161-173.

Davelaar, E. J., Gosher-Gottstein, Y. Ashkenazi, A., Haarman, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112,* 3-42.

Daily, L. Z., Lovett, M. C., & Reder, M. L. (2001). Modeling individual differences in working memory performance: a source activation account. *Cognitive Science, 25,* 315-353.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegner, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *PNAS, 106,* 6802-7.

Horn, D., & Usher, M. (1992). Oscillatory model of short term memory. In J. E. Moody, S. J., Hanson, & R. P. Lippmann (Eds.), *Advances in Neural Processing and Information Systems Vol. 4* (pp. 125-132). Morgan and Kaufmann.

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

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139,* 39–50.

Jensen, O., & Lisman, J. E., (1998). An oscillatory short-term memory buffer model can account for data on the Sternberg task. *Journal of Neuroscience, 18,* 10688-99.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition, 24,* 103-106.

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.-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.

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, 13,* 766-785.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review, 18,* 324-330.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428,* 751-754.

Usher, M., Cohen, J. D, Haarmann, H., & Horn D. (2001). Neural mechanism for the magical number 4: Competitive interactions and nonlinear oscillation. *Behavioral and Brain Sciences, 24,* 151-152.