Dynamical Origin of the Effective Storage Capacity in the Brain's Working Memory

Christian Bick[1,2,3,*] and Mikhail I. Rabinovich[1,†]

[1]*BioCircuits Institute, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0402, USA*
[2]*Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0112, USA*
[3]*Network Dynamics Group, Max Planck Institute for Dynamics and Self-Organization (MPIDS), 37073 Göttingen, Germany*

(Received 3 June 2009; published 19 November 2009)

The capacity of working memory (WM), a short-term buffer for information in the brain, is limited. We suggest a model for sequential WM that is based upon winnerless competition amongst representations of available informational items. Analytical results for the underlying mathematical model relate WM capacity and relative lateral inhibition in the corresponding neural network. This implies an upper bound for WM capacity, which is, under reasonable neurobiological assumptions, close to the ''magical number seven.''

DOI: 10.1103/PhysRevLett.103.218101 PACS numbers: 87.18.Sn, 05.45.−a, 87.10.−e, 87.18.Hf

Introduction.—Working memory (WM) is the ability to transiently hold and manipulate several items in the mind, which are involved in the immediate information processing or actions such as thinking, planning, and eventually, behavioral output. Tasks involving WM include, for example, remembering a sequence of statements that we recently heard in a lecture or following driving directions to an unfamiliar place. Language, as sequential activity, is also based on WM. However, the capacity of WM is limited and that is one reason why the metaphor ''blackboard of the mind'' became popular to describe WM. Its capacity is defined as the number of items that can be recalled after a WM task. It varies amongst different individuals and also with age. Numerous studies have led to the generally accepted point of view that the effective capacity ranges between three to seven items [1–4]. This limit has coined the term ''magical number seven'' [5] in conjunction with WM.

In this Letter, we propose a model for sequential WM (SWM) that is based on winnerless competition (WLC) [6] between informational items. Items in this SWM are represented by saddle fixed points and the mnemonic recall by a trajectory in a stable heteroclinic channel (SHC) [7,8]. In contrast to attractor dynamics, the transient itself reflects the sequential memory. The main result we report is that, under certain simplifying assumptions, there is an upper bound on the number of items that can be stored in this SWM model when implemented by inhibitory-coupled neuronal clusters. This upper bound depends upon the relative strength of lateral inhibitory cell connections in the corresponding neural network. Assuming that connection strengths are sampled from uniform distributions and that the relative connection strengths amongst the inhibitory-coupled units cannot exceed an order of magnitude, the bound for the number of items is about seven. This is remarkable because, although the model itself has not yet been directly experimentally confirmed, its neural dynamics exhibit the very same inherent bound for SWM capacity as given by the ''magical number.''

Model.—SWM dynamics is separated into two stages: storage of the sensory information and its retrieval. Storage means initiation of a specific pattern in the phase space of the corresponding dynamical system by both sensory input and the contents of WM. We hypothesize that WLC between different informational items is the main mechanism for retrieval in SWM. WLC is a widely known phenomenon in systems involving more than two interacting agents that satisfy a relationship similar to the popular rock-paper-scissors game or the voting paradox [9,10]. The participants of such a process can become winners periodically, or, especially when the number of participants is more than three, the process can be noncyclic. As a generic dynamical phenomenon, which is rare in simple systems yet common in complex ones, the stimulus dependent sequence of switching among informational items can provide a concise and constructive formulation of WM dynamics.

In terms of nonlinear dynamics, we identify informational items with saddle fixed points (or, more generally, with saddle sets) in the phase space of the corresponding dynamical system. The information the memory holds is now stored in the structure of the phase space by representing a sequence in this SWM by a heteroclinic sequence or, more precisely, by a SHC. Consecutive temporary winners are represented by the separatrices connecting the corresponding saddles. Different channels joining different saddles result in different contents of the SWM as illustrated in Fig. 1(a). In order to get robust and reproducible recall dynamics, dissipativity of the saddle equilibria, i.e., the condition that the phase space volume around the saddles is compressed rather than stretched, and structural stability are key properties for memory performance. The mnemonic recall is now given by a trajectory in the SHC, a replay of the competition. The trajectory spends time in the vicinity of every temporary winner, leading a macroscopic ''switching behavior.''

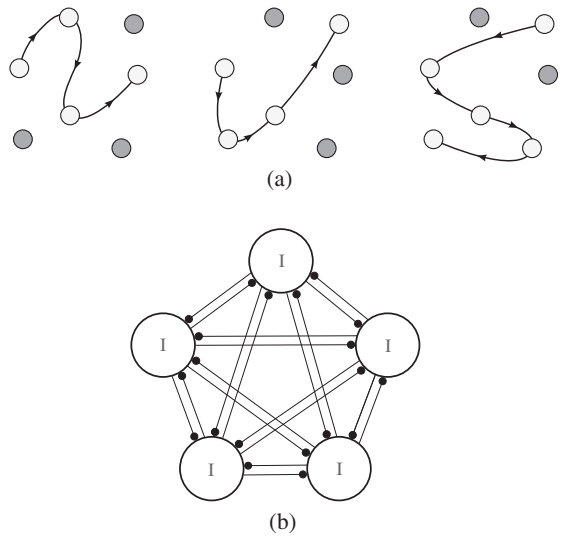

FIG. 1. (a) Different transient trajectories correspond to different temporary winners. The light circles depict informational items that are activated sequentially, i.e., are stored in SWM, whereas the dark circles represent inactive items. (b) A SWM network with five interacting nodes. The inhibitory-coupled, competing clusters are denoted by I and the lines with dots depict inhibitory connections. Real neurobiological settings are much larger.

On a neuronal level, such competition can be implemented with inhibitory connections between interneurons in the prefrontal cortex. Because of them, neurons can temporarily suppress the other competing neurons to emerge as the temporary winner. Sequential switching of activity patterns of groups of prefrontal interneurons thus is the result of WLC of inhibitory interacting neural clusters. The model network we propose consists of overlapping neural clusters with both inhibitory and excitatory connections. The excitatory coupling is important for supporting the activity patterns for a finite amount of time. The all-to-all inhibitory coupling, on the other hand, is the core dynamical mechanism for sequential competitive switching; cf. Fig. 1(b).

The storage capacity of this SWM can now be defined as the maximal length of such a sequence that can be supported by the phase space under reasonable neurobiological constraints. Intuitively, the amount of competition should increase with the number of WM items. Thus, long chains of switching are not robust against perturbations and do not produce the memorized sequence as it is observed in nature (see, for example, [2,11]). Based on the existence and stability conditions, we are looking for a dependence of the parameters of the neuronal network on the number of steps (items) in the sequence.

It is important to note that the time scale of WM (the time a sequence can be held on-line) and its capacity (the number of items that can be recalled) are determined by different neurophysiological mechanisms [12–14]. In our model capacity (the length of the sequence) and time scale (the time in which the structure of the phase space can be adapted) are two different features, so we are able to focus on the dynamical origin of SWM capacity without discussing the time scale. On-line manipulation of the contents of SWM means modulation of the phase space. Therefore, our model predicts that this feature is also independent of the capacity limit.

Analytical results.—Since generalized Lotka-Volterra equations describe the dynamics of the inhibitory interacting agents of the network, they are the basis for our mathematical description of the SWM network. The model is given by the following set of ordinary differential equations:

$$\dot{a}_i = a_i\Big[\sigma_i(M,S) - \Big(a_i + \sum_{j\neq i}\varrho_{ij}(M,S)a_j\Big)\Big], \quad (1)$$

for $i = 1,\ldots,N$, where $a_i(t) \geq 0$ describes the neuronal activity of an informational mode or firing pattern of a neural cluster that represents the ith informational item to be stored, $\varrho_{ij}(M,S) \geq 0$ describes the inhibitory connections between the ith cluster and the jth cluster, and $\sigma_i(M,S)$ represents the level of self-excitation.

Note that the excitation and connection strengths are functions of two variables. In real biological networks, the connection strengths are context dependent and could be changed dynamically by top-down (M) or bottom-up (S) inputs. Bottom-up inputs are described by S and correspond to sensory input or early feature detection mechanisms, associated with early attentional selection [15]. Conversely, M plays the role of top-down modulatory effects. These could be associated with mnemonic retrieval or late attentional processing [16,17]. These dependencies can be related to the storage and on-line manipulation mechanism because for a recall we assume that the connection strengths are "set," leading to the metastable dynamics described below.

The mathematical representation of WLC dynamics that lead to switching behavior in the considered network is given by a SHC. A heteroclinic sequence is a sequence of saddles, i.e., metastable equilibria in the phase space, so that each one is a saddle point with a one-dimensional unstable manifold, and consecutive saddles in the sequence are joined by a heteroclinic orbit. Here $A_i := (0,\ldots,0,\sigma_i,0,\ldots,0)$, with σ_i being the ith entry of the vector A_i, are the corresponding equilibria and let $\iota = (i_1,\ldots,i_K)$ denote the ordered set of indices corresponding to this sequence of equilibria of length K. A heteroclinic sequence is called stable if the saddles are dissipative; i.e., the saddle values ν_i satisfy $\nu(A_i) = -\mathrm{Re}\lambda_{2i}/\mathrm{Re}\lambda_{1i} > 1$, where λ_{ji} are the eigenvalues of the linearization around A_i ordered monotonically decreasing by real part. Then the system (1) contains a stable heteroclinic channel, which is a suitable ε-neighborhood V of the stable heteroclinic sequence so trajectories that have their initial point in the vicinity of A_1 stay in V for finite time T [18] as illustrated in Fig. 2.

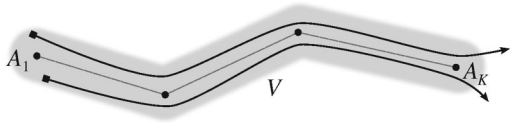

FIG. 2. Trajectories in a SHC.

In a neurobiological setting it makes sense to consider networks in which the connection strengths ϱ_{ij} are drawn randomly. This situation was studied in [19] and the main result can be summarized in the following theorem.

Theorem 1 [Bick and Rabinovich (2009)] Let ι be any open sequence of equilibria of length $K \in \mathbb{N}$ and suppose that the entries of σ corresponding to indices in ι form a Fibonacci sequence, i.e., $\sigma_{i_1} = a$, $\sigma_{i_2} = ac$, $\sigma_{i_{k+1}} = \sigma_{i_k} + \sigma_{i_{k-1}}$, with $\sigma_{i_k} < 2\sigma_{i_{k+1}}$, $k = 1, \ldots, K$ and the remaining entries are sufficiently small. Then there are intervals J, J' on the positive real axis and $\mu > 0$ large enough such that for any $c \in J$, $a > 0$ the grossly organized system (1) with a random coupling matrix, where $\varrho_{i_{k-1}i_k}$, $\varrho_{i_{k+1}i_k}$ are picked randomly from the interval J' and ϱ_{ii_k}, $i \notin \{i_{k-1}, i_k, i_{k+1}\}$ are picked randomly from the interval (μ, ∞), has a structurally stable SHC in $V(\iota, \varepsilon)$ for every sufficiently small $\varepsilon > 0$. ■

The context of the theorem is essential for the results below. Its proof is based on the reinvestigation of the conditions for the existence of a SHC given in [18]. There, under the assumption of a random stimulus σ, coupled inequalities for the coupling strengths were derived. The theorem above follows from uncoupling these inequalities in such a way that, instead of σ being random and ϱ_{ij} depending on σ, the stimulus σ is fixed and the ϱ_{ij} are now random within certain bounds. For a quantification of these bounds and a detailed proof see [19].

The proof of the theorem is based on the assumption of a Fibonacci sequence. However, for the following results, not the Fibonacci sequence itself but more its geometric growth is important. This growth condition reflects the increasing need of excitation to support long sequences since longer sequences lead to the activation of more inhibitory connections. We want to quantify this observation. The maximal range of the parameter c can be defined as the set

$$A := \left\{ c > 0 \;\middle|\; \begin{array}{c} \sigma_{i_k} < 2\sigma_{i_{k+1}} \quad \text{and} \\ \max\limits_{k \le K} \frac{\sigma_{i_{k+1}}}{\sigma_{i_k}} - \frac{1}{2} < \min\limits_{k \le K} \frac{\sigma_{i_{k+1}}}{\sigma_{i_k}} \end{array} \right\}.$$

We have $A \neq \varnothing$. Furthermore with

$$C'(K, c) := \min_{k \le K-1} \left(\frac{\sigma_{i_{k+1}}}{\sigma_{i_k}} \right)$$

we obtain the bound

$$\sup J' \le C(K) = \max_{c \in A} C'(K, c).$$

Because of the convergence of the ratio of consecutive members of a Fibonacci sequence, both C' and C are bounded. For the parameter μ we have

$$\mu \ge \mu_s(K, c) = \max_{i \in I, k \le K} \left(\frac{\sigma_i}{\sigma_{i_k}} \right) + 1.$$

We obtain a lower bound for relative lateral connection strengths for elements drawn from the different uniform distributions by dividing the minimum of the larger one by the maximum of the smaller one. Therefore, we can define a ''scaling function''

$$\Phi(K) := \frac{\min\limits_{c \in A} \mu_s(K, c)}{\max\limits_{c \in A} C'(K, c)} \le \frac{\mu_s(K, c)}{\sup J'} \le \frac{\varrho_{ii_k}}{\varrho_{i_{k-1}i_k}} \qquad (2)$$

with $i \notin \{i_{k-1}, i_k, i_{k+1}\}$, which now only depends on the length of the sequence, i.e., the parameter K. This function is plotted in Fig. 3 and grows like a geometric sequence with increasing K.

In other words, the ratio of randomly selected inhibitory connections, relative to lateral connections that impose order on the heteroclinic sequence with normalized self-inhibition, is lower bounded by $\Phi(K)$. Given an upper bound on the relative connection strengths, this implies that there is an upper bound on the length of the sequence. In a neurobiological setting, quantities are always bounded and experimental data [20–22] suggests that an upper bound for Φ of about 10–20 is realistic. A plot of $\Phi(K)$, depicted in Fig. 3, reveals that the resulting limit of the sequence length adequately matches the ''magical number seven'' for working memory.

Discussion.—The model of SWM, using heteroclinic sequences of metastable states (items) to store information, gives robust and reproducible transient dynamics for memory recall. It also reflects Cowan's idea of activation of stored memory patterns [14] that is related to experimental data. Recently, experimental evidence for the dependence

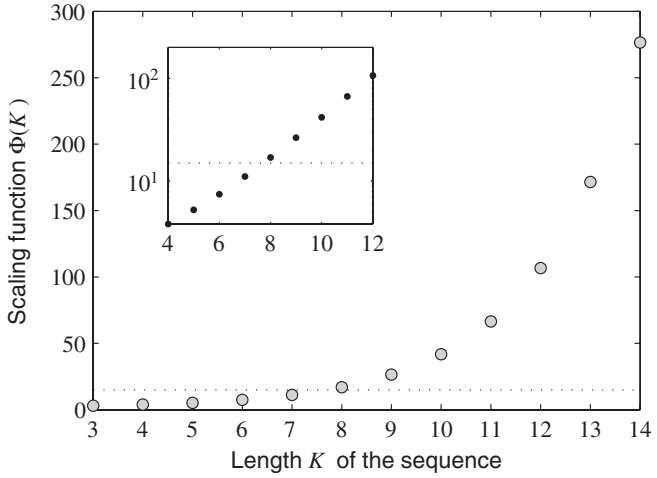

FIG. 3. Relationship between the relative connection strengths and the number of saddles. For a sequence length of over ten the function $\Phi(K)$ rises very quickly to give minimal ratios that are neurobiologically unrealistic. The inline plot shows the same data on a log scale. The dotted line depicts $y = 15$ in both graphs.

of the capacity limitation for visual WM on the level of inhibition was produced by Edin *et al.* [4]. They confirmed the predictions of their computational model, consisting of excitatory and inhibitory elements, with a functional magnetic resonance imaging (FMRI) study.

We considered the important case of one-dimensional unstable separatrices of metastable states. In complex systems there are areas in control parameter space with two-, three-, and even higher-dimensional unstable separatrices [23]. However, close to the bifurcation boundaries, the positive Lyapunov exponents satisfy the inequality $\ldots, \lambda_3, \lambda_2 \ll \lambda_1$. Although it has not yet been proved, it is intuitively clear that in that case a trajectory will follow the direction of λ_1 with a high probability; therefore, something resembling a stable heteroclinic channel will still exist. In fact, it is widely accepted that, because of learning, neural systems operate in control parameter regions close to bifurcation boundaries.

It has been reported that memory performance decreases with increasing memory load [24,25]. In the framework of our model, this can be explained by two independent phenomena. The first effect is based on the bifurcations discussed above. Higher-dimensional separatrices lead to uncertainty in the itinerary, therefore resulting in poorer SWM performance. Second, informational items are represented in the phase space of the corresponding dynamical model by saddle equilibria. During memory recall the trajectory ''passes through'' the vicinity of the stored items. The closer the minimal distance to the saddle point, the longer the period of time the pattern is active, thus resulting in well-separated firing patterns. On the other hand, if the distance between saddles and trajectory becomes larger, the recall is ''blurred out,'' and firing patterns overlap or ''interact,'' therefore resulting in poorer memory performance.

We did not discuss the storage mechanism that is responsible for setting up the phase space to store a memory sequence. It is reasonable to hypothesize a ''competitive learning rule'' where an initial competition, in which the winners are able to set connection strengths, is responsible for the setup so that a ''replay'' (recall) becomes possible. Moreover, we assume that the informational items themselves correspond to specific patterns of activity, i.e., saddle points in the phase space. Therefore, these representations must have been encoded in the network at some point in order to be associated with different stimuli. It is important to note that this kind of learning is different from the process that is responsible for the storage of sequences in the phase space. However, the bifurcations of high-dimensional dynamical systems, which must be responsible for that kind of encoding, are still poorly understood.

In conclusion, although the model has not yet been directly verified experimentally, the remarkable result of the very same inherent bound for SWM capacity, as given by the ''magical number,'' gives a possible explanation of SWM dynamics.

The authors would like to thank the anonymous referees for valuable hints and constructive suggestions to improve the presentation of the results. M. I. R. would like to acknowledge support by grant ONR N00014-07-1-0741. C. B. would like to thank Gert Cauwenberghs for making this project possible.

*bick@nld.ds.mpg.de
†mrabinovich@ucsd.edu

[1] H. L. Swanson, Dev. Psychol. **35**, 986 (1999).
[2] K. Oberauer and R. Kliegl, J. Mem. Lang. **55**, 601 (2006).
[3] J. N. Rouder, R. D. Morey, N. Cowan, C. E. Zwilling, C. C. Morey, and M. S. Pratte, Proc. Natl. Acad. Sci. U.S.A. **105**, 5975 (2008).
[4] F. Edin, T. Klingberg, P. Johansson, F. McNab, J. Tegnér, and A. Compte, Proc. Natl. Acad. Sci. U.S.A. **106**, 6802 (2009).
[5] G. Miller, Psychol. Rev. **63**, 81 (1956).
[6] M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. I. Abarbanel, and G. Laurent, Phys. Rev. Lett. **87**, 068102 (2001).
[7] M. Rabinovich, R. Huerta, and G. Laurent, Science **321**, 48 (2008).
[8] M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, PLoS Comp. Biol. **4**, e1000072 (2008).
[9] *Classics of Social Choice*, edited by I. McLean, A. B. Urken, and F. Hewitt (The University of Michigan, Ann Arbor, MI, 1995), pp. 81–89.
[10] D. G. Saari, *Basic Geometry of Voting* (Springer-Verlag, Berlin, 1995).
[11] K. Oberauer, Mem. Cognit. **37**, 346 (2009).
[12] G. Mongillo, O. Barak, and M. Tsodyks, Science **319**, 1543 (2008).
[13] A. V. Egorov, B. N. Hamam, E. Fransen, M. E. Hasselmo, and A. A. Alonso, Nature (London) **420**, 173 (2002).
[14] N. Cowan, *Attention and Memory: An Integrated Framework* (Oxford University Press, New York, 1998).
[15] G. R. Mangun, Psychophysiology **32**, 4 (1995).
[16] N. Carlisle and G. Woodman, J. Vision **9**, 180 (2009).
[17] T. P. Zanto and A. Gazzaley, J. Neurosci. **29**, 3059 (2009).
[18] V. S. Afraimovich, V. P. Zhigulin, and M. I. Rabinovich, Chaos **14**, 1123 (2004).
[19] C. Bick and M. I. Rabinovich, ''On the Occurrence of Stable Heteroclinic Channels in Lotka-Volterra Models,'' Dyn. Syst. (to be published).
[20] R. Miles, J. Physiol. **431**, 659 (1990).
[21] H. Markram, J. Lübke, M. Frotscher, A. Roth, and B. Sakmann, J. Physiol. **500**, 409 (1997).
[22] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, PLoS Biol. **3**, e68 (2005).
[23] V. Afraimovich, I. Tristan, R. Huerta, and M. I. Rabinovich, Chaos **18**, 043103 (2008).
[24] P. M. Bays and M. Husain, Science **321**, 851 (2008).
[25] R. Sapkota, S. Pardhan, A. Tavassoli, and I. van der Linde, Ophthalmic Physiolog. Opt. **28**, 99 (2008).