Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neural Networks 19 (2006) 375–387
Neural Networks
www.elsevier.com/locate/neunet

A composite neural network model for perseveration and distractibility in the Wisconsin card sorting test

Gülay B. Kaplan [a,*], Neslihan S. Şengör [b], Hakan Gürvit [c], İbrahim Genç [a], Cüneyt Güzeliş [d]

[a] Information Technologies Research Institute, TÜBİTAK-Marmara Research Center, Gebze, 41470 Kocaeli, Turkey
[b] Faculty of Electrical Electronics Engineering, İstanbul Technical University, Maslak, 80626 İstanbul, Turkey
[c] Behavioral Neurology and Movement Disorders Unit, Department of Neurology, İstanbul Faculty of Medicine, İstanbul University, Çapa, 34490 İstanbul, Turkey
[d] Department of Electrical and Electronics Eng., Engineering Faculty, Dokuz Eylül University, Buca, 35160 İzmir, Turkey

Received 14 December 2003; accepted 11 August 2005

Abstract

A composite artificial neural network model is proposed to simulate the performance of the Wisconsin Card Sorting Test. The Wisconsin Card Sorting Test is a test of executive functions where prefrontal deficits are matched to some quantitative measures such as percentage of perseverative errors and number of failures to maintain set. In this work, the proposed model is used to simulate the performances of healthy subjects and patients with prefrontal involvement particularly on these measures. The model is designed in such a way that one of the subsystems, namely, the Hopfield network, serves as the working memory and the other, the Hamming block, as the hypothesis generator. The results show that the proposed relatively simple model is capable of simulating the wide range of the performances of both normal subjects and prefrontal patients on the Wisconsin Card Sorting Test. While lowering the Hamming distance in the Hamming block gave rise to progressively more perseverative responses, changing the threshold vector of the Hopfield network resulted in more set maintenance failures. The former manipulation disrupts the abstraction or mental flexibility and the latter sustained attention or perseverance both of which are the major functions of the prefrontal system.

Keywords: Computational modeling; Prefrontal cortex; Executive functions; Wisconsin card sorting test; Perseveration; Distractibility; Hopfield network; Hamming network

1. Introduction

Research in the field of computational modeling is a part of the attempts to unreveal the computations performed by the brain. Computer modeling not only allows hypothesis testing, but it is also a powerful mean of generating new and original hypotheses (Eysenck & Keane, 1990; Rugg, 1997; Stillings, Weisler, Chase, Feinstein, Garfield and Rissland, 1995). A computational model can provide novel sources of insight to the brain function by providing alternative explanations for the observed phenomena. Production systems constitute one approach where cognitive structures can be modeled in a computationally powerful way (Kimberg & Farah, 1993; Polk, Simen, Lewis, & Freedman, 2002). These models are good at explaining complex behavior, but they are insufficient to generate behavioral deficits. The other approach, namely, artificial neural networks (ANN), is much better for the simulation of behavioral deficits (Berdia & Metz, 1998; Cardoso, 1991; Cardoso & Parks, 1998; Changeux & Dehaene, 2000; Cohen, Dunbar, & McClelland, 1990; Frank, Loughry, & O'Reilly, 2001; Sohn & Gaudiot, 1992). The proposed model is an example of the latter approach.

The prefrontal cortex (PFC) is the area of the brain that is associated with high level, 'executive' processes which are needed for goal-directed behavior (Fuster, 1997). Frontal lobes and particularly their prefrontal sub-sectors undergo a striking expansion along the course of mammalian evolution and frontal lobes reach an unmatched size in the primate and particularly in the human, occupying approximately 30% of the cerebral hemispheres of the human brain (Goldman-Rakic, 1987). The prefrontal cortex is the most highly interconnected

Abbreviations aCG, anterior cingulate; ANN, artificial neural network; dl-PF, dorsolateral prefrontal; FMS, failure to maintain set; LLS, learning to learn score; OF, orbitofrontal; PFC, prefrontal cortex; RC, response card vector; **v**, feature vector; **T**, threshold vector; **TC**, template card vector; **W**, weight matrix; WCST, wisconsin card sorting test.

* Corresponding author. Tel.: +90 262 641 2300x4722; fax: +90 262 646 3187.

E-mail address: gulay.kaplan@bte.mam.gov.tr (G.B. Kaplan).

0893-6080/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2005.08.015

region of the cortex. It sends and receives projections from all the sensory-motor association cortices, limbic structures and many subcortical structures such as the thalamus and the striatum which are components of the parallel fronto-striatal circuits (Miller & Asaad 2002; Alexander, Crutcher & Delong 1990; Alexander & Crutcher 1990). As a result, the PFC has access to a wide variety of refined information about the external physical world and the internal milieu of the organism and holds a unique position for orchestrating often conflicting demands of external reality and internal drives which are essential for voluntary goal-directed behavior. There are various views on how many subdivisions the PFC might have, but the traditional subdivision follows cytoarchitectonical lines: a neocortical dorsolateral prefrontal (dl-PF) region and a mesocortical orbito-frontal one (Mesulam, 2000b). Anatomical connectivity patterns and behavioral data suggest that prefrontal cortex can be subdivided into three discrete circuits and thus add a third anterior cingulate circuit (Alexander et al.. 1990). All three regions form parallel circuits that are closed loops interconnected with subcortical structures, i.e., discrete sub-sectors of caudate nucleus and thalamus. These circuits subserve different functional specializations: dl-PF circuit is responsible for executive functions, OF circuit is for comportment or socially mediated behavior and aCG circuit mediates motivation. Dysfunctions of the individual circuits give rise to recognizable discrete clinical pictures that are dysexecutive syndrome for dl-PF, disinhibition syndrome (socially inappropriate behavior and impulsivity) for OF and apathy for aCG (Mega & Cummings 1994; Tekin & Cummings 2002). However, in clinical practice, it is usual to see prefrontal patients with mixed features from two or all three syndromes paralleling the extent of the lesion encroaching the boundaries of those circuits.

The term executive functions is used to refer higher-order abilities such as abstraction, judgement, reasoning, decision-making, planning, anticipation and resistance to interference of the inappropriate stimuli (Fuster, 1997; Goldman-Rakic, 1987; Shallice, 1988; Stuss & Benson 1986). The term working memory has been proposed to describe a mechanism for the temporary active or on-line maintenance and manipulation of information (Baddeley, 1986). According to Fuster (1997) prefrontal cortex is critical in the temporal organization of behavior in order to reach a predetermined goal.

Several neuropsychological tests have been developed so far in order to explain the functions associated with the prefrontal system. Among the most widely used, one can mention the WAIS-R similarities sub-test and Raven's Progressive Matrices for abstract reasoning, Tower of Hanoi or Tower of London tests for planning, the Trail Making Test for mental set shifting, and Stroop and Go-No Go tests for response inhibition to name a few (Boone, 1999). The Wisconsin Card Sorting Test (WCST) is the most commonly used test to asses set-shifting abilities in humans. In this article only the WCST is considered and the aim is to propose a computational model, which exploits the PFC processes during the WCST.

The WCST requires abstraction, hypothesis testing, mental flexibility and the ability to maintain or alter responses according to positive or negative reinforcement. Humans with prefrontal damage cannot perform this task since they are unable to flexibly adapt their behavior to the regularly changing demands of the test.

Our main aim is to propose a computational model based on the hypothesis that the major function of the prefrontal cortex is to liberate the organism from stimulus-bound behavior and enable it to act upon the stimulus in a context-dependent fashion (Cohen et al., 1990; Lhermitte, Pillon, & Serdaru, 1986; Mesulam, 1986, 1998). Thus, the more developed the prefrontal cortex is, the more flexible is the organism in its selection of alternative actions to a given stimulus, depending on the contextual cues. On the other hand, a developed prefrontal cortex endows the organism with the capacity of working memory that is on-line maintenance and manipulation of information during the temporal unfolding of the goal-directed behavior.

We suggest that, the normal and abnormal performance on the WCST can be effectively simulated when the major symptoms associated with prefrontal damage, namely, perseveration as an index of poor flexibility and distractibility as an index of poor attention or working memory as revealed by the WCST (Greve, Williams, Haas, Littell, & Reinso, 1996; Greve, Ingram, & Bianchini, 1998; Koren, Seidman, Harrison, Lyons, Kremen and Caplan, 1998), are taken into consideration using artificial neural networks. In this work, we propose a computational model for this purpose. While the model is developed different subsystems are proposed for simulating perseveration and distractibility, during analysis simulation results, it was observed that the subsystem responsible for distractibility also affected perseveration. Thus, the model exploits how positive effect of distractibility on perseveration can occur.

2. The Wisconsin card sorting test

Although there are different versions of the test, the commonly used one is the original test that was introduced by Milner (Heaton, Chelune, Talley, Kay, & Curtis, 1993; Milner, 1963). In this version, there are two identical 64-card decks and four reference or template cards. The cards are different with respect to their three features, namely, color, shape, and number of items on each card. Each feature can take four different values (color, red, green, blue, yellow; shape, triangle, star, cross, circle; number, 1, 2, 3, 4). So there are 64 (4^3) different cards. The template cards are those which are different from each other in all three features and the rest are called response cards. The template cards are shown in Fig. 1 and remain throughout the task. During the test, these template cards are aligned horizontally in front of the subject, and the subject is given the first deck and asked to open the cards one by one. The task is to match each card of the deck with one of the four template cards. As shown in Fig. 1, the first template card matches in color with the response card, the second with number etc. After every single matching of the subject, the experimenter responds only 'correct' or 'false' according to the category considered by experimenter during that particular

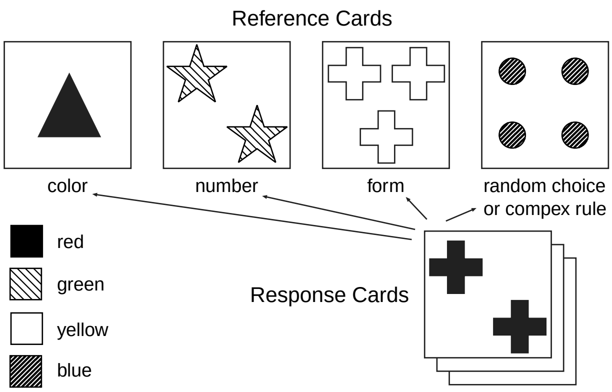

Fig. 1. Material used in the WCST (Dehaene, & Changeux, 1991).

time interval and makes no other comment. The subject must infer the correct sorting feature based on the feedback from the experimenter. The subject can match a card in four ways: either according to one of the three features or to an 'other' fourth obscure rule which the subject considers in his/her mind, one which does not fit with any of the three features as illustrated in Fig. 1. Unbeknownst to the subject, the criterion changes after certain number of correct matchings according to a predetermined order. In the used version of the test, the criterion is changed after 10 correct responses. The order of experimenter's criterion is color, shape, number, and repetition of this 3-tuple. The test is terminated when the subject completes six categories or the entire 128 cards are exhausted (Heaton et al., 1993).

Then a scoring procedure is applied to measure the performance of the subject. In WCST the deficits are reflected principally as lower number of categories and increased percentage of perseverative errors and increased number of set maintenance failures. A perseverative response in the WCST is the insistency on matching to the previous correct category, although it no longer yields a 'correct' response. Poor planning, poor abstraction, deficient mental flexibility, hence poor hypothesis generation and testing would prevent the subject deciphering the valid rule and he/she would be unable to respond appropriately to the negative feedback after the rule has changed and would tend to insist on the previous response pattern. Thus, the patient in accordance with the severity of his/her difficulties would complete fewer categories and commit high percentage of perseverative errors (Levine & Prueitt, 1989; Nelson, 1976). On the other hand, problems of sustained attention and perseverence would cause distractibility, inability to inhibit inappropriate response tendencies and thus interfere with the working memory function. It is well known that prefrontal damage causes a noisy internal milieu. Knight and Grabowecky (2000) cite a study in which prefrontal patients displayed an impaired performance in detecting a subsequent target when an irrelevant stimulus is presented in the intervening delay period and in contrast, they improved their performance when no such stimulus was present during the delay period. The distractibility would in turn be reflected in subject's difficulty in keeping up responding according to the valid rule, hence commit 'failure to maintain set' (FMS). The FMS is arbitrarily defined in the WCST as the inability to maintain the rule and reach the criterion of completing a category despite a minimum of five consecutive correct matchings (Heaton et al., 1993). Every such breaking of the rule is counted as one FMS. In other words, the subject displays evidence that the new rule has already been conceptualized by responding the minimum number of consecutive correct choices but cannot sustain his/her attention and gets distracted before the category has been completed.

In the factor analytic studies of the WCST, the above mentioned assumption that the perseverations and FMS reflect executive and attentional functions, respectively, had already been shown (Greve et al., 1996; Greve et al., 1998; Koren et al., 1998). The first factor, that is perseveration accounted for 58–70% for the variance and was always the best discriminator between the patients (mostly schizophrenics) and the normal subjects. Greve et al. (1998) further tested specifically the hypothesis that the consistently found second factor, that is the FMS does indeed reflect attentional function and found confirming evidence.

One can argue that discrete lesions of the prefrontal system, involving different complex fronto-striatal circuits, result in more or less identifiable error patterns in WCST. The dysexecutive patient secondary to dorsolateral prefrontal circuit involvement is more prone to achieve fewer categories and, score a higher percentage of perseverative responses. Heaton singles out the perseveration score as the best WCST measure for the presence of frontal involvement (Heaton et al., 1993). Inverse correlation between dl-PF metabolism and perseverative response scores were also demonstrated in a positron emission tomography (PET) study (Lombardi, Andreason, Sirocco, Rio, & Gross, 1999). One might think that impulsive, inattentive, disinhibited patient secondary to OF circuit involvement would tend to break the rule in FMS fashion. In the experimental literature a consistent neuroanatomical substrate is not evident for the FMS. In one study it was demonstrated that scores on a smell identification test correlated inversely only with FMS score in patients with schizophrenia (Stedman & Clair, 1998). Since OF includes primary olfactory cortex, this finding can be taken as an evidence of the relationship of the FMS score and the dysfunctional OF system. Shulman, Fiez, Corbetta, Buckner, Miezin and Raichle (1997), demonstrated that the activity of the OF in PET decreased while a novel task is progressively learned and as the performance of the subjects increased. That is, more attentional resources are allocated to inhibit probable distractors when the situation is novel, thus OF takes over and when the skills to perform the task are acquired OF may be left to rest. Finally, the increasing evidence implicates a dysfunctional OF for a primary disorder of attention such as attention deficit hyperactivity disorder (ADHD) (Shepard, Bradshaw, Purcell, & Pantelis, 1999). Alternatively, a hemispheric asymmetry could be proposed to account for the dysfunctional neural substrates of perseveration and distractibility. Using PET Raichle, Fiez, Videen, MacLeod, Pardo and Fox (1994) demonstrated that a novel verb generation task activates left prefrontal cortex and this activity decreases with practice and as the performance increases. Raichle (2000, p. 1316) stated this study further suggests that this finding is consistent with 'the broad view that prefrontal cortex is active when new rules need to be learned and older ones rejected'. This is exactly what happens when a sorting rule has been changed in WCST. Conversely, it is well-known that the right hemisphere is dominant for directed spatial attention as evidenced by numerous experimental data, along with left hemi-neglect stroke patients. Furthermore, disorders of global attention that are acute confusional states do occur after focal lesions, but mainly secondarily to those involving the heteromodal cortices of the right hemisphere (for a review of the both disorders of attention see Mesulam (2000a)). Finally, a more widespread involvement of the both circuits would result in a combination of both type of errors. As suggested by the above mentioned factor analytic studies, we have singled out perseveration and distractibility in the WCST and specifically focused on them in this model, which will be explained in the next section.

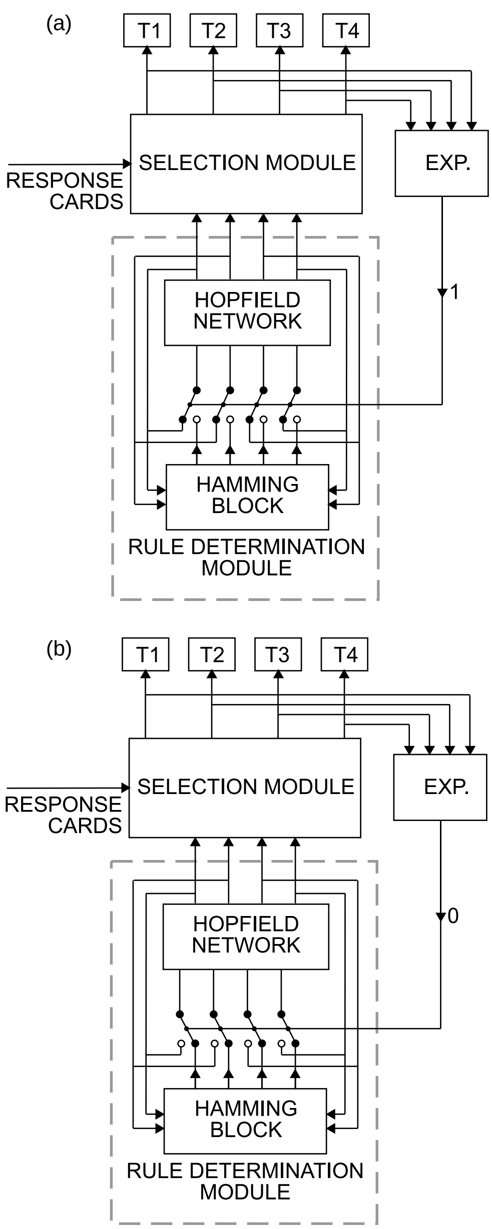

Fig. 2. (a) When the exp.'s response is 'correct' (b) When the exp.'s response is 'false' The modes of the proposed composite neural network model for WCST.

3. Proposed composite neural network model

In the proposed model, it is assumed that the subject's behavior is completed in two phases for each test item. We propose that, these phases consist of two hypothetical tasks. These tasks are supposed to be fulfilled by the subjects to perform the test successfully. In the first phase, the selection rule, which the subject proposes to select a template card, is determined. In the next phase, the template card, which has the same value with the response card, is chosen based on the rule determined in the first phase. These two phases are realized by 'rule determination' and 'selection' modules in the model, as shown in Fig. 2.

The first phase, i.e. the process of determining the selection rule, is performed considering the experimenter's response to the previous decision of the subject. If the experimenter's response is 'correct', the subject should maintain the current rule, if it is 'false', the alternatives should be considered. This corresponds to temporary activation or on-line maintenance of information, namely, the working memory mechanism (Miller & Cohen, 2001; O'Reilly, Noelle, Braver, & Cohen, 2002). The maintenance of the valid rule in the proposed model is accomplished by the Hopfield network (see Fig. 3(a)). On the other hand, if the experimenter's response was false then the Hamming block takes over (see Fig. 3(b)) and offers to the Hopfield network an alternative selection rule by setting up its initial state. This change in the selection rule is a consequence of a negative reinforcement, while keeping the same rule is a response to a positive reinforcement. Thus, negative reinforcement triggers the Hamming block. The effect of reinforcement is on the hypothesis generator. In Leven & Levine (1987), reinforcement activates bias nodes, which take part in attention

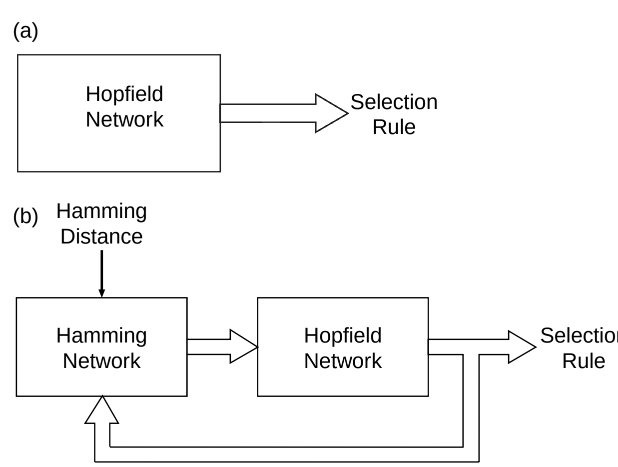

Fig. 3. (a) When the exp.'s response is 'correct' (b) When the exp.'s response is 'false' The block diagrams corresponding to the activated part of the model.

Table 1
Coded features

Features					
Color		Shape		Number	
Red	1 0 0 0	Triangle	1 0 0 0	1	1 0 0 0
Green	0 1 0 0	Star	0 1 0 0	2	0 1 0 0
Yellow	0 0 1 0	Cross	0 0 1 0	3	0 0 1 0
Blue	0 0 0 1	Circle	0 0 0 1	4	0 0 0 1

drawing, and in Levine and Prueitt (1989) past and present values of reinforcement affect novelty detection.

In order to construct this model, first of all cards of WCST are represented as binary vectors of 12 elements. Since the cards are discriminated by their three features (color, shape, and number) and each feature could have four different values, each feature is coded in four digits and they are concatenated to form the card vector. This concatenation is done in order of color, shape, and number. The corresponding codes for features are depicted in Table 1.

The selection rule is the term used to define the criterion that the subject holds as the valid rule during sorting. The subjects may think the selection rule as one of the discriminative features (color, number, shape) and sort accordingly. This is an unambiguous response in WCST terminology. However, since 40 cards share two attributes with the same reference card, e.g. the card with two red triangles is similar to the first reference card (‘one red triangle’) in both color and shape, if sorted likewise the experimenter would not figure out according to which of the two criteria it had been sorted and therefore it is called an ambiguous response. The selection criterion of the subject is represented by a vector **v**, which is named as the rule vector. This vector is the output of the rule determination module and at the same time it is the input vector of the selection module (Figs. 2 and 3). Although the cards are discriminated by three features, i.e. color, shape, and number, **v** has four components (first three for the discriminative features and fourth for the ‘other’, i.e. an idiosyncratic sorting which does not fit any of the three criteria). Subjects often imagine some complex rule decipherable only perhaps by some kind of arithmetic and sort likewise. Therefore, there are a total of 16 possible binary vectors of dimension four. The vectors [1 1 1 1] and [0 0 0 0] are meaningless for the proposed model and, therefore, excluded: the former vector signifies that the subject either simultaneously considers color, shape, number, and the ‘other’ as the rule and the latter vector is its exact opposite, that is, simultaneous deferral of all four possible alternatives, both of which are virtually impossible. The remaining 14 vectors are selection rule vectors given in Table 2. Three of them refer to discriminative feature vectors, which are named base feature vectors and the fourth for the ‘other’. The other 10 are the vectors for all of the remaining possible ‘ambiguous’ selections and called as multi-feature selection rule vectors. All of these selection rule vectors are stored in the Hamming block to provide alternative rules to be taken into account by the Hopfield network. Thus, the Hamming block is composed of 14 neurons.

When the experimenter’s response is ‘false’, the subject has to consider a new selection rule (Figs. 2(b) and 3(b)). In this work, this process is triggered by the Hamming block (explained in the Appendix A), which proposes a new selection rule to the Hopfield network. Thus, this new selection rule vector is derived from the current one according to a pre-specified Hamming distance, which is defined as the number of different bits between two binary vectors. When the Hamming distance is high it means the subject can think of different rules. When the Hamming distance is low, the subject is attracted to the previous rule. The Hamming block not only includes a Hamming network, but also a mechanism that is capable of generating vectors from the current rule vector, within a predefined Hamming distance. As there can be more than one vector, which have the same Hamming distance to the current one, one of them should be selected randomly. Six different values, that is 0, 1, 2, 3, 1–2, and 2–3 are used as Hamming distances during the simulations. In 1–2 and 2–3 cases these two numbers are mixed with equal probability. Five experiments with the first number and five experiments with the second number were done. The higher is the Hamming distance value, the more is the probability of generating a different criterion vector.

Table 2
The meaning of the rule vectors

Rule vector	Meaning
[1 0 0 0]	Color
[0 1 0 0]	Shape
[0 0 1 0]	Number
[0 0 0 1]	Other
[1 1 0 0]	Color and shape
[1 0 1 0]	Color and number
[1 0 0 1]	Color and other
[0 1 1 0]	Shape and number
[0 1 0 1]	Shape and other
[0 0 1 1]	Number and other
[1 1 1 0]	Color and shape and number
[1 1 0 1]	Color and shape and other
[1 0 1 0]	Color and number and other
[0 1 1 1]	Shape and number and other

The Hopfield network determines the subject’s selection rule, therefore, it is composed of four neurons. Since the criterion vector is binary, discrete Hopfield network is utilized. For simulating the performance of a successful subject, it is expected that the output of the Hopfield network is to be one of the fundamental feature vectors and also this rule has to be maintained when the experimenter’s response is ‘correct’ (Figs. 2(a) and 3(a)). Therefore, the weight matrix and threshold vector of the Hopfield network are determined to give fundamental feature vectors as the fixed points of the Hopfield network. When the experimenter’s response is ‘false’, the Hamming block takes its turn and proposes an alternative selection rule to the Hopfield network as an initial condition. Thus, the Hamming block controls the Hopfield network by taking precedence when the external cues indicate an obligation to shift the set (Figs. 2(b) and 3(b)). Thus, different rules can be rapidly activated through this control mechanism.

The above re-initialization procedure of the defined Hopfield network renders it different from the ordinary usage of Hopfield network (Hopfield & Tank, 1985). Hopfield networks are used conventionally as associative memory and in solving optimization problems. In both applications once an initial condition is given either a pattern in the memory or a sub-optimal solution of optimization problem is obtained. Here, the non-linear convergent dynamics of the Hopfield network is adopted to generate different selection rules according to different initial conditions (Hirsh, 1989). Thus the re-initialization corresponds to the rapid updating of different strategies, rules or goals by simply changing the activation state of a set of neurons (O'Reilly et al., 2002). If the Hamming block proposes more than one feature (i.e. the rule vector has at least two '1'), then the Hopfield network must choose one of them as the selection rule. If such is the case, the aim of the Hopfield network is to reduce the alternative rules to one which corresponds to a fundamental feature. This requirement is also taken into account while the weight matrix and threshold vector of the Hopfield network are determined. It is possible to have other weight matrix and threshold vector values that give fundamental vectors as fixed points. The well-known procedure to obtain weight matrix $\mathbf{W}$ and threshold vector $\mathbf{T}$ is the Hebbian learning rule (Haykin, 1994). However, the Hebbian rule may give rise to undesired fixed points, that are multi-feature selection rules. In this work, the weights and threshold values are determined by solving the algebraic equations obtained from the Hopfield's difference equations given in Appendix A.

The equations are solved by setting an appropriate fundamental feature vector as the next state for each possible state. With this procedure of obtaining $\mathbf{W}$ and $\mathbf{T}$, there is no possibility of ending up in undesired fixed points. If undesired fixed points are obtained, reducing alternative rules to one would be broken and this would cause an inappropriate selection of response cards. The process of reducing the alternative rules can be broken modifying the threshold vector of the Hopfield network. So the subject tries to use more than one feature instead of a single fundamental feature. We used two different threshold vectors, namely, $\mathbf{T}_1$ and $\mathbf{T}_2$ for the modification purpose. There are other choices of threshold values to obtain similar results. In fact, we tried several vectors but concluded that including results with $\mathbf{T}_1$ and $\mathbf{T}_2$ was sufficiently representative for a disordered state. The weight matrix and threshold vectors of the Hopfield network are given in Appendix A.

The 'selection' module we use is similar to Winner-Take-All (WTA) network which is a well known network for clustering purposes (Zurada, 1992). In the conventional WTA network, the inner product of input and weight vectors is used, whereas in this module, the weighted norm in Eq. (1) suitable for selection purpose is being used. Another difference from the conventional WTA is that there is no learning phase; instead of cluster centers being determined during a learning phase, the template cards are used as fixed cluster centers. Each neuron of the selection module matches to a template card, in turn. So the outputs of selection module correspond to the subject's choice for current response card as shown in Fig. 2. The inputs of selection unit neurons are the response card vector and determined selection rule vector which is the output of the rule determination module. The four neurons in the selection module compute a weighted norm of the difference between the response card vector and template card vector as given in Eq. (1). The output of the selection network is the index of reference card with minimum distance defined by the weighted norm.

$$\|\mathbf{RC}-\mathbf{TC}^i\|_w = \sum_{k=1}^{3} v_k \left[\sum_{j=1}^{4} \left|RC_{(k-1)4+j} - TC^i_{(k-1)4+j}\right|\right] + v_4 \sum_{k=1}^{4}\left[2-\sum_{j=1}^{4}\left|RC_{(k-1)4+j} - TC^i_{(k-1)4+j}\right|\right] \tag{1}$$

where $\mathbf{TC}^i$ is used for ith template card vector and $\mathbf{RC}$ for response card vector.

The model was designed so that the Hopfield network serves as the working memory, as it is indeed responsible for the on-line maintenance of the valid rule, and the Hamming block as the planner and hypothesis generator, providing abilities which are indispensable to shift a mental set. It is expected that the model, as it is, should function as an healthy subject (i.e. six categories, no FMS's and an acceptable percentage of perseverative errors, which is under 10%). In the normative study 150 normal subjects commited a mean of 11.8% of perseverative errors with a standard deviation (SD) of 7.1 and 0.8 FMS (SD: 1.3) (Heaton et al., 1993). Tempering the threshold vector of the Hopfield network should simulate the lesioned working memory and therefore the model should yield primarily higher FMS's. However, distractable state can also increase the probability of a perseverative response, since the rule-breaking erroneous response will be a perseverative one, with 50% probability. Moreover, the tendency to perseverate in the previously valid rule is itself a strong source of distraction. On the other hand, tempering the Hamming block, that is, decreasing the Hamming distance value should simulate an unoptimal hypothesis generator by lessening its capacity to propose different selection rules and, therefore, the model should primarily yield higher percentage of perseverative errors, but also fewer completed categories, whereas not necessarily FMS's. Finally, tempering both networks should simulate a more widespread prefrontal lesion and the model should yield both higher perseverative responses and higher number of FMS's along with fewer, even no completed categories.

The WCST material and application of the test are simulated in a software environment using MATLAB® and the results are scored according to the WCST scoring procedure. In the simulation, response card vectors are ordered as in the original test of Milner (Heaton et al., 1993). Our personal experience with the test indicates that a big majority of the normal and clinical subjects starts by matching first response card which is a 'one green triangle' with the template card that is 'one red triangle' thus matching to both shape and number features, so committing an ambiguous false response. Only a minority does otherwise, that is choosing 'two green stars', thus a correct

Table 3
The simulation results of WCST

C	HD	HT	#CR	#CC	PE%	FMS
1	3	T	64.6±3.5	6.0±0	9.9±2.8	0±0
2	3	T_1	73.7±6.1	1.1±0.5	19.7±2.9	3.2±1.6
3	3	T_2	66.2±8.4	0.3±0.5	26.5±6.4	2.4±1.8
4	2	T	67.6±3.7	6.0±0	14.2±4.6	0±0
5	2	T_1	66.3±9.1	0.4±0.5	24.6±4.9	2.6±2.1
6	2	T_2	66.7±9.9	0.5±0.7	27.9±8.6	2.0±1.5
7	1	T	39.4±0.5	1.0±0	67.8±0.5	0±0
8	1	T_1	58.8±6.9	1.1±0.4	32.4±6.2	1.1±0.6
9	1	T_2	61.9±8.3	0±0	27.7±6.8	2.4±1.2
10	0	T	32.8±2.5	0.1±0.3	73.0±1.5	0±0
11	0	T_1	31.5±4.7	0±0	38.6±4.8	0±0
12	0	T_2	30.4±4.1	0±0	39.4±3.5	0±0
13	3-2	T	64±1.8	6.0±0	9.5±2.1	0±0
14	3-2	T_1	75.9±6.1	1.6±1.0	17.7±3.4	2.7±1.3
15	3-2	T_2	72.6±7.9	0.8±0.8	25.9±6.2	3.1±1.7
16	2-1	T	70.7±5.5	6.0±0	17.9±3.7	0±0
17	2-1	T_1	73.2±6.3	1.3±0.5	21.2±5.0	2.4±1.2
18	2-1	T_2	59.8±6.3	0.1±0.3	29.1±6.9	2.0±0.9

C, Condition; HD, Hamming distance; HT, Hopfield threshold; CR, number of correct responses; CC, Completed categories; PE%, perseverative error percentage; FMS, failure to maintain set score; numbers after ± are standard deviations (SD's).

response (the predetermined first rule is color). The first response is assigned as shape and number with the probability of 90% and color 10%. Only for the first card, the selection modules run and select a template card. After the experimenter's response is generated, the Hamming block or the Hopfield network takes turn. If the experimenter's response is 'false', the Hamming block generates alternative selection rule and transfers it to the Hopfield network and the new selection rule is determined by the Hopfield network for the next card. If the experimenter's response is 'correct', only the Hopfield network runs. Having determined the selection rule, the second card is applied and selection module chooses a template card by using Eq. (1). When there is more than one '1' in the criterion vector, there may be numerous template cards minimizing the weighted norm in Eq. (1), so one is favored over the others randomly. These processes are repeated until six categories are completed or all the 128 cards are exhausted.

As described previously in this section using a combination of six different Hamming distance values (3, 3–2, 2, 2–1, 1 and 0) and three different Hopfield vectors ($\mathbf{T}$, $\mathbf{T}_1$ and $\mathbf{T}_2$), a total of 18 condition are created. There are 10 trials in each condition. The means and standard deviations of the four chosen WCST

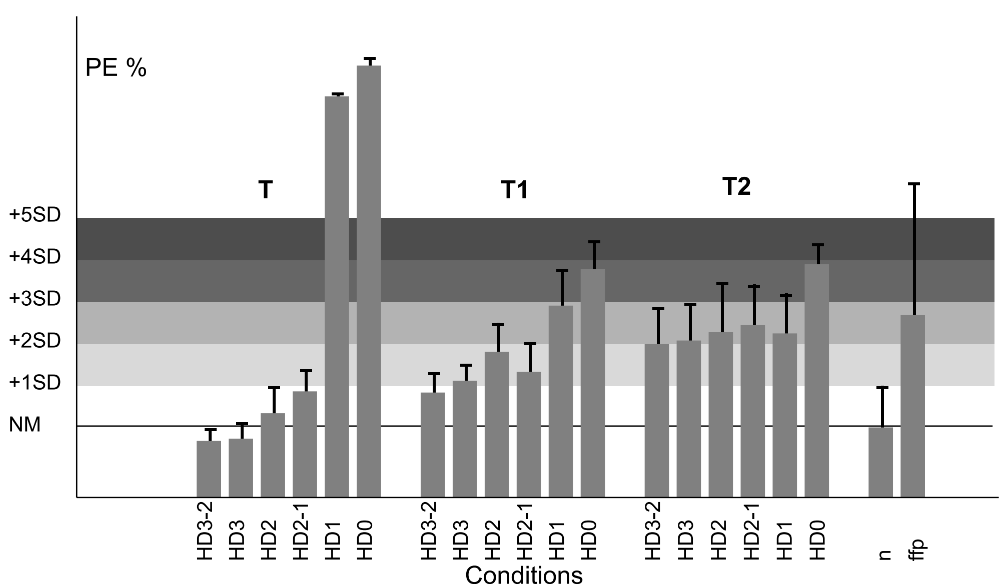

Fig. 4. Perseverative error percentages with different Hamming distance values and under different Hopfield network thresholds. PE%, perseverative error percentage; n, normal subjects ($n=150$); and ffp, all focal frontal patients ($n=73$) in Heaton et al.'s (1993) normative sample. Columns and bars represent the means and standard deviations (SD's) respectively. NM: normative mean (11.8 PE%). The graying rows represent the number of SD's from the mean of the normals (+1 SD=18.9 PE%, +2 SD=26 PE%, +3 SD=33.1 PE%, +4 SD=40.2 PE%, +5 SD=47.3 PE%).

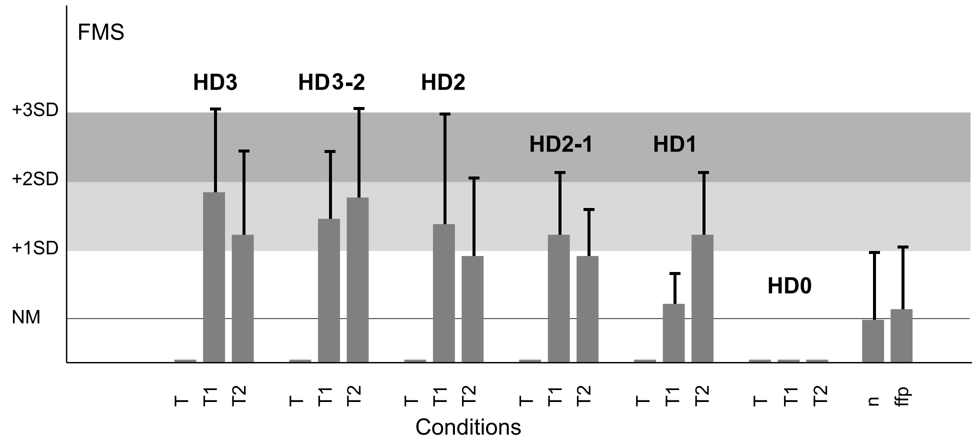

Fig. 5. Failure to maintain set scores under different Hopfield network thresholds and with different Hamming distance values. FMS, failure to maintain set score; *n*, normal subjects ($n=150$); and ffp, all focal frontal patients ($n=73$) in Heaton et al.'s (1993) normative sample. Columns and bars represent the means and standard deviations (SD's), respectively. NM: normative mean (0.8 FMS). The graying rows represent the number of SD's from the mean of the normals (+1 SD=2.1 FMS, +2 SD=3.4 FMS, +3 SD=4.7 FMS).

measures (number of correct responses, number of completed categories, perseverative error percentage, and FMS score) for each of the 18 conditions are shown in Table 3.

4. Simulation results

The performance of the system in 18 different conditions for four WCST measures are shown in Table 3. In Figs. 4 and 5 the means and SD's of the performances of the 18 conditions in two measures the WCST are plotted with those of the 150 normal subjects and 73 all focal frontal patients' of Heaton et al. (1993) normative sample. The graying color scale in the figures represents the number of SD's over the normative mean of that particular measure.

As it can be followed from Table 3, when the Hamming distance value is kept fixed and the threshold vector is changed from $\mathbf{T}$ to $\mathbf{T}_1$ or $\mathbf{T}_2$, ability of set maintenance decreases as reflected by increased FMS's, which in turn increases perseverative response tendency with the exception of C7 and C10, where the Hamming distances are 1 and 0, respectively, and where the system is most rigid in terms of producing an alternative in response to negative feedback. On the other hand, keeping the threshold vector fixed and decreasing the Hamming distance values decrease flexibility as reflected by increased perseverations. Nevertheless, this increase stays within normal limits of Heaton et al. (1993) normative sample even in C16, where the Hamming distance is as low as being an equal combination of 2 and 1.

Of interest is the fact that the 'pathological' rigidity becomes evident abruptly and severely when the flexibility of the system is as limited as in Hamming distance 1. In condition 7 (C7) and condition 10 (C10), where the Hamming distances are 1 and 0, respectively, the so-called hypothesis generator clearly fails. In these circumstances the model fails to produce alternatives to changing situations and insists on its old habits despite continuous negative feedback. The model is otherwise quite consistent at least responding habitually with no FMS's, thus sometimes achieving at least a single category. In C10 the model keeps up with its first ever response, so since the probability of matching the first card to color category is 10%, there is only one trial in which at least a single category is completed. Neuropsycholgists are accustomed to this kind of 'consistency' of the prefrontal patients, even not infrequently witnessing their overt dissociation of consciously admitting the change of the rule, but still failing to inhibit the magnetic attraction of perseverative response: often a prefrontal patient would say 'this is also false' aloud, while responding perseveratively.

Interestingly, distractibility has an almost positive effect on perseveration when the Hamming distance is 1. In conditions 8 (C8) and 9 (C9) due to the effects of $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively, the percentages of perseverative errors significantly highly decrease as compared C7 and the total numbers of correct responses increase. Despite this effort it does not result in true improvement in both cases, since the number of completed categories remains the same in C8 or even further decrease in C9 as compared to C7. Conceivably, modification of the Hopfield network distracts the habitual responses, that is, the obstinate sticking of the model with its perseverative responses and pushes it to choose a random alternative at least for a while, but since distraction also prevents the perseverance, the model commits a set maintenance failure before completing a category.

In conditions 11 (C11) and 12 (C12) the results are similar to those of C8 and C9, the only difference being no committed FMS's in these conditions. Apparently, with Hamming distance 0, the distraction effect of either $\mathbf{T}_1$ or $\mathbf{T}_2$ becomes so much that the model cannot even persist in five consecutive correct responses, which is the arbitrary rule for a set maintenance failure to be rated as an FMS.

5. Discussion and conclusion

In this work, a model for simulating WCST task performance, a conventional test of prefrontal functioning, is developed by using neural networks. These functions, unlike the ‘non-frontal’ cognitive functions such as language, spatial attention or memory, which are more or less clearly delineated in terms of both neuropsychology and neuroanatomy, are still conceptually vague, being composed of numerous, often overlapping terms and definitions. For those non-frontal cognitive functions large-scale distributed networks have been proposed with epicenters in clearly defined anatomical sites, connectivity patterns and relative behavioral specializations, supported with an extensive clinical as well as newly emerging functional magnetic resonance imaging (fMRI) data (Mesulam, 1981, 1990, 1998, 2000b; Morecraft, Geula, & Mesulam, 1993). Nevertheless, the distinction that the dl-PF is executive and OF is for response inhibition, as Fuster has suggested, is not only useful clinically, but also we suggest that it helps to model WCST in a simple but comprehensive way. A left hemispheric hypothesis generator (Wolford, Miller, & Gazzaniga, 2000) and a right-hemispheric sustained attention modulator (Posner, 1994) hypothesis are also compatible with our model. Successful performance in WCST is conceived as on-line maintenance and perseverance on the rule as long as it is reinforced and appropriate rejection of the old rule and shift the set when a new rule is introduced. Perseverance is largely dependent on an intact working memory system that is actively safeguarded against interference, disruption of which gives rise to a state that is called distractibility. In our model, changing the threshold vector of the Hopfield network, which is stood for the working memory module, gave rise to distractibility as reflected by the increasing number of FMS’s committed. On the other hand, the inefficiency of the Hamming block, which served as hypothesis generator, was reflected in the increasing percentage of perseverative errors. Note that perseverative error percentage is highest in conditions 7 and 10, where the Hamming distances are 1 and 0, respectively, and the slope in conditions $\mathbf{T}$ is quite steep, as shown in Fig. 4. Sometimes a perseverative sequence may represent a kind of perseverance, even if faulty, as evidenced by the mid-range performance of conditions $\mathbf{T}_1$ and $\mathbf{T}_2$ where the system is so distractible that it commits set maintenance failures and thus sometimes cannot maintain even a perseverative sequence.

As shown in, Fig. 5 FMS score abruptly rises when the Hopfield network threshold is tempered. Note that in conditions 10–12, where the Hamming distance is 0, there are no FMS’s. This is probably because in these conditions the system becomes so distractible and helpless in terms of alternative hypothesis generation that the responses are completely haphazard and it cannot persevere even in five consecutive correct responses, which is the arbitrary definition of FMS in WCST.

We think that our model has certain advantages over the previous attempts to model WCST task performance (Amos, 2000; Berdia & Metz, 1998; Dehaene & Changeux, 1991; Kaplan, Genç, & Güzeliş, 2001; Leven & Levine, 1987; Monchi, Taylor, & Dagher, 2000), one being its capability to simulate the performances of a wide array of normal subjects and prefrontal patients on a continuum of most flexible to gradually most rigid subjects, despite its relatively simple structure. The network proposed by Leven & Levine (1987) is based on adaptive resonance theory network (ART). Here the main idea is to develop a model that is capable of displaying the behavior of normal and PFC lesioned subjects from the perspective of perseveration. In the model the experimenter’s response has a weighted effect which is provided by a gain parameter and by changing the value of it, one healthy and lesioned subject is simulated. The healthy subject is capable of completing five categories and the lesioned one is not capable of changing the feature from color to another. So the simulation results are quite limited. There is no possibility of observing the mechanisms that create distractibility and different perseveration levels.

In the work of Dehaene & Changeux (1991) the main idea is to develop a tool to provide functional analysis of WCST and this is done by using six different mechanisms that are further modelled by a specific neural network. In order to obtain the results of PFC-damaged patients, some mechanisms are disrupted. While a theoretical analysis is provided, their model is too complex to simulate the behavior of the subjects during WCST. Among the given simulation results only perseveration, which is a diagnostic measure of PFC damage, is modelled.

Monchi et al. (2000) developed a computational model based on the basal ganglia-thalamocortical loops for WCST. Different modifications are applied to the model in order to investigate the working memory deficits observed in patient with Schizophrenia and Parkinson’s disease. Simulation results show the activities of the circuits for normal and patient with Schizophrenia and Parkinson’s disease. Although it is shown that tempering model parameters causes impairment of decision making process, how this impairment is reflected in measures of WCST such as perseverative error percentage or FMS is not accounted.

Another model, where only perseverative responses are considered, is Amos’s (2000) model. Amos’s (2000) model also considers the neuroanatomic loops from the frontal cortex through the basal ganglia and thalamus. He proposed a neural network model considering the connections of the loops and possible effects of dopamine release in them. Simulation results show the perseverative responses of the patient with Schizophrenia, Parkinson’s and Huntington’s diseases for WCST. While in this work different modules of corticobasal loops are considered each of these modules has a similar formula for the activation and two parameters in these formulas, namely, bias and gain are modified to obtain the results. Our approach is completely different, we proposed subtasks and carrying out these subtasks are formed using different structures.

Berdia & Metz (1998) designed a neural network model that is able to simulate performance of normals and schizophrenics on the WCST. Although the model motivated by biological considerations, they did not take into account the functions of PFC sub-circuits. In the model there are category neurons, which laterally inhibit each other for one category to fire at a time. There is a feedback loop, which takes part in the learning phase of a rule, whatever the experimenter’s response is. If it is correct the efficacy of synaptic weight becomes stronger and vice versa. The simulation of the different error types is modelled by the synaptic instability, which occurs due to the noise parameters. They performed huge number of simulations to determine the effects of noise parameters. They considered performance indices related to errors and categories in order to show the effect of noise.

In another article on modelling PFC in WCST, (Kaplan et al., 2001), a mixed procedure of Milner's and Nelson's versions of WCST is used (Nelson, 1976). This relatively simple system consists of a decision network, which determines the selection rule, and an action network, which selects one of the template cards by using the selected rule. The action network is a Winner-Take-All network while the decision network is a one-hidden-layer perceptron. During the training of the one-hidden-layer perceptron, the input consists of subject's current and two past selection criteria and the experimenter's response while desired output is the selection criterion for the next response card. The desired outputs are only discriminative features. During the simulations, after the first randomly selected criterion, the action network selects one of the template cards and the experimenter's response is generated, then the trained one-hidden-layer perceptron takes action and determines the criterion for the next card. What is understood from perseveration is the prefrontal patient's stronger emphasis on his/her past selections instead of the experimenter's response as imposed by Leven & Levine's (1987) work. The principal strategy for the simulation of the patient's this behavior in this model was to attenuate the effect of experimenter's response on selection process. Modifying the weights between hidden neurons and experimenter's response and the biases of these neurons, perseverative responses were obtained. The modifications of the weights are in a direction to lessen the effect of the experimenter's response. The simulation results are given only for one healthy and one lesioned subject. While the healthy subject achieves all category shiftings, the lesioned one could not change his/her selection rule from color as in Leven & Levine (1987). Although the system is simple when compared the other models, its modeling of PFC in WCST is limited with perseverations of the single lesioned subject.

In conclusion, our model is able to simulate various good and poor performances on the WCST that can be obtained by normal subjects and prefrontal patients. One can draw an analogy between the components of our model, namely, 'the hypothesis generator' the Hamming block and 'the working memory module' the Hopfield network and the sub-sectors of the prefrontal cortex, dl-PF and OF, respectively. As mentioned before, in the dysexecutive syndrome associated with dl-PF damage poor abstractions and planning give rise to stimulus-bound, perseverative behavior, and in the OF syndrome disinhibition and impulsivity result in distractibility and inability to resist interference, both of which render the goal-directed behavior inefficient, if not impossible (Fuster, 1997). In our model, the normal performance becomes more and more like a prefrontal patient's, on the one hand as the alternatives of the Hamming block are progressively restricted and rendered stimulus-bound, and on the other hand as the threshold vector of the Hopfield network is modified and rendered distractible. This is in accordance with the hypothesis that the essential function of the human prefrontal cortex is the goal-directed behavior which is enabled by the flexible modification of the components of the behavior as the contextual cues signal negative feedback, and which is safeguarded against the interference of inappropriate stimuli as it is temporally unfolded in the process of reaching its goal.

However, we cannot claim to have simulated every aspect of the WCST task performance. One missing aspect is the 'other' response. Some of the examinees tend to commit the 'other' response, sometimes because of the completely haphazard nature of the response tendencies of a prefrontal patient, yet others do so, because an otherwise normals subject fails to decipher one of the simple fundamental criteria, but instead theorizes on some complex rule that glimpses only in his/her mind. The latter performance usually results in a reduced number of completed categories, while perseverative error percentage remains in the acceptable level. There is not such a gray zone in our results of completed categories. The performance is at either of the two extremes, excellent-good (i.e. 6 or 5 categories) or null-bad (i.e. 0 or 1 category). We think that the failure to represent this mid-zone is probably due to absence of the 'other' responses. In our model neither the Hamming block, ever considered proposing an 'other' response, nor a distracted Hopfield network broke the rule in order to try randomly an 'other' response. A second missing aspect might be the learning efficiency of a normal subject during the WCST performance as reflected by 'learning to learn score' (LLS). Having noticed the change of the rule after a certain number of correct responses normal subjects usually increase their efficiency and find the subsequent rule with less and less trials and errors. This is usually achieved through the second part of the test (2nd set of color-form-number sequence) and reflected in a higher positive LLS value. Our model fails to display this learning ability.

The above mentioned lacking aspects of the model are due to concern of modeling only the most illustrative measures of WCST, namely, perseveration and distractibility. We think that improving our model by appropriate procedures to incorporate the simulation of those lacking aspects would enhance its capability and worth the effort of further study.

Appendix A

Hopfield network is a recurrent neural network which has a dynamical behavior giving rise to more than one fixed point (Haykin, 1994; Hopfield & Tank, 1985; Zurada, 1992). These fixed points correspond to different representations of objects obtained from the same structure and their interpretation depends on the context. This dynamical behavior is not only due to non-linear system equations but it is also due to some constraints on weight matrix which is either obtained by design constraints or Hebbian learning rule (Haykin, 1994; Zurada, 1992). In either case these weights can be considered as values for 'hard-wired' structure, as once they are determined they are fixed values (Monchi et al., 2000).

In this article, the dynamic behavior of the Hopfield network is obtained by the following difference equation:

$$\mathbf{x}_{\mathrm{Hop}}(k+1) = \mathbf{W}_{\mathrm{Hop}}\mathbf{y}_{\mathrm{Hop}}(k) - \mathbf{T}_{\mathrm{Hop}} \tag{2}$$

$$\mathbf{y}_{\mathrm{Hop}}(k) = f(\mathbf{x}_{\mathrm{Hop}}(k)) \tag{3}$$

In this system $\mathbf{x}_{\mathrm{Hop}}$, $\mathbf{y}_{\mathrm{Hop}} \in \{-1,1\}^4$, $\mathbf{W}_{\mathrm{Hop}} \in \Re^{4x4}$, $\mathbf{T}_{\mathrm{Hop}} \in \Re^4$ and the non-linear function $\mathbf{f}_{\mathrm{Hop}}(\cdot)$ as follows:

$$\mathbf{f}_{\mathrm{Hop}}(v) = \begin{cases} 1 & \text{if } v > 0 \\ -1 & \text{if } v < 0 \\ v & \text{if } v = 0 \end{cases} \tag{4}$$

As the state space is finite, it is well-known that such dynamics either ends in a limit cycle or a fixed point (Goles-Chacc, Fogelmann-Soulie, & Pellegrin, 1985). The dynamical system given by Eqs. (2–4) has only three fixed points corresponding to 'color', 'shape', and 'number' when the weight matrix $\mathbf{W}_{\mathrm{Hop}}$ and the threshold vector $\mathbf{T}_{\mathrm{Hop}}$ are taken as follows:

$$\mathbf{W}_{\mathrm{Hop}} = \begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \qquad \mathbf{T}_{\mathrm{Hop}} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -4 \end{bmatrix}$$

While solving the Eq. (2) $\mathbf{x}$(k) values are updated asynchronously. As the Hopfield network with bipolar activation function is used, coding of rule vectors are different then Table 2 in Section 3. In the coding now there are 1 and -1 and these correspond to 0 and 1, respectively.

This choice of weight matrix and threshold vector is not the unique, but with this weight matrix and the threshold vector 14 of the meaningful initial conditions end in one of the three fixed points either in one or two steps.

In order to obtain other choices of rules which would correspond to different fixed points, it is enough to change threshold vector. So, when $\mathbf{T}_1 = [1\ \ 1\ \ 1\ \ 1]^{\mathrm{T}}$ is used instead of $\mathbf{T}$ four fixed points but this time corresponding to 'color', 'color, number, other', 'shape, number, other' and 'color, shape, other' are obtained. The other choice is $\mathbf{T}_2 = [2\ \ 2\ \ 1\ \ 2]^{\mathrm{T}}$ which gives two fixed points 'shape, number, other' and 'color, shape, other'. It is possible to obtain different rules just by changing threshold vector, which corresponds to changing the place of fixed points in the finite space $\{-1,1\}^4$.

Hamming network is a feed-forward neural network, which in most applications used with MAX-NET (Bose & Liang, 1996) to determine the vectors that are different from each other according to maximum Hamming distance.

It obeys the following equation:

$$\mathbf{y}_{\mathrm{H}} = \frac{1}{2}\mathbf{W}_{\mathrm{H}}\mathbf{u}_{\mathrm{H}} + \mathbf{T}_{\mathrm{H}} \tag{5}$$

$\mathbf{y}_{\mathrm{H}} \in Z^{14}$, $\mathbf{u}_{\mathrm{H}} \in \{-1, 1\}^4$, $\mathbf{W}_{\mathrm{H}} \in \{-1, 1\}^{14x4}$, $\mathbf{T}_{\mathrm{H}} \in \Re^{14}$

The weight matrix $\mathbf{W}_H$ and the threshold vector $\mathbf{T}_H$ are taken as follows (Mehrotra, Mohan, & Ranka, 1997):

$$\mathbf{W}_{\mathrm{H}} = \begin{bmatrix} \mathbf{u}_1^T \\ \bullet \\ \bullet \\ \mathbf{u}_p^T \end{bmatrix}, \qquad \mathbf{T}_{\mathrm{H}} = \begin{bmatrix} -\dfrac{n}{2} \\ \bullet \\ \bullet \\ -\dfrac{n}{2} \end{bmatrix} \tag{6}$$

Here, $p=14$ and $n=4$. $\mathbf{u}_{\mathrm{H}}$ corresponds to rule vectors obtained from the Hopfield network. However, the rule vectors are transformed to unipolar codes as exploited in Table 2 in Section 3 while running the codes. The rows of the $\mathbf{W}_{\mathrm{H}}$ matrix correspond to 14 meaningful rule vectors. $\mathbf{y}_{\mathrm{H}}$ gives the Hamming distance between the current rule vector obtained from the Hopfield network and the stored 14 rule vectors.

In order to explain how the model works when considering a flexible subject, the Hopfield threshold vector and the Hamming distance are taken as $\mathbf{T}$ and 3, respectively. With these parameter values the model runs for the first two steps and the results are given in the following for each step. The first response card is 'one green triangle' and it is represented by the card vector [0010 0010 1000]. The selection module choose the third template card and this is expected since the first selection rule is [0 1 1 0] with 90% probability. As the answer of the experimenter is 'false' the Hamming block produces [1 0 0 0] which is at the Hamming distance 3 to [0 1 1 0], and this vector corresponds to the initial condition of the Hopfield vector for the next step. As this is the fixed point of the Hopfield network its output is same as its initial condition. This selection rule will be kept till the experimenter's response is 'false'.

Another example is given where $\mathbf{T}_1$ is the Hopfield threshold vector and the Hamming distance is 1. In this case, just like the above example the first selection rule is again [0 1 1 0]. To the experimenter's 'false' response the Hamming block produces [1 1 1 0] for the second card which is 'four red cross'. The vector [1 1 1 0] corresponds to 'color, shape, number' and it is at 1 Hamming distance to [0 1 1 0]. The Hopfield network using this vector as its input produces [1 0 1 1] and the model selects the fourth template card which means that number is taken as the determinative feature. Again the answer is 'false', the Hamming block takes turn and the same process continues till the sixth response card. As for the seventh card the determinative feature is 'color', the experimenter's response is 'correct'. The model produces consecutive 10 'correct' responses since the 'color' is the fixed point of the Hopfield network with this threshold vector. At the end of the test only one category is completed with 55 correct responses.

References

Alexander, G., & Crutcher, M. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*, 266–271.

Alexander, G., Crutcher, M., & Delong, M. (1990). Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculumotor, prefrontal and limbic functions. *Progress in Brain Research*, *85*, 119–146.

Amos, A. (2000). Computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, *12*, 505–519.

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Berdia, S., & Metz, J. (1998). An artificial neural network stimulating performance of normal subjects and schizophrenics on the Wisconsin Card Sorting Test. *Artificial Intelligence in Medicine*, *13*, 123–138.

Boone, K. (1999). Neuropsychological assesment of executive functions. In B. Miller, & J. Cummings (Eds.), *The human frontal lobes functions and disorders* (pp. 247–260). NewYork: The Guilford Press.

Bose, N., & Liang, P. (1996). *Neural network fundamentals with graphs, algorithms and applications*. New York: Mc Graw-Hill.

Cardoso, J. (1991). Revisiting the Tower of Hanoi. *AI Expert*, 49–53.

Cardoso, J., & Parks, R. (1998). Neural network modeling of executive functioning with the Tower of Hanoi test in frontal lobe-lesioned patients. In R. Parks, D. Levine, & D. Long (Eds.), *Fundamentals of neural network modeling* (pp. 209–231). Massachusetts: The MIT Press.

Changeux, J., & Dehaene, S. (2000). Hierarchical neuronal modeling cognitive functions: From synaptic transmission to the Tower of London. *International Journal of Psychophysiology*, *35*, 179–187.

Cohen, J., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Dehaene, S., & Changeux, J. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*, 62–79.

Eysenck, M., & Keane, M. (1990). *Cognitive psychology*. London: Psychology Press.

Frank, M., Loughry, B., & O'Reilly, R. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective and Behavioral Neuroscience* , 137–160.

Fuster, J. (1997). *The prefrontal cortex*. Philadelphia: Lippincott–Raven.

Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representiall memory. In F. Plum (Ed.), *Handbook of physiology: The nervous system* (Vol. 5) (pp. 343–417). Bethesda, MD: American Physiological Society.

Goles-Chacc, E., Fogelmann-Soulie, F., & Pellegrin, D. (1985). Decreasing energy functions as a tool for studying threshold networks. *Discrete Applied Mathematics*, *12*, 201–277.

Greve, K., Ingram, F., & Bianchini, K. (1998). Latent structure of Wisconsin Card Sorting Test performance in a clinical sample. *Archives of Clinical Neuropsychology*, *13*, 597–609.

Greve, K., Williams, M., Haas, W., Littell, R., & Reinso, C. (1996). The role of attention in Wisconsin Card Sorting Test performance. *Archives of Clinical Neuropsychology*, *11*, 215–222.

Haykin, S. (1994). *Neural networks a comprehensive foundation*. New York: Macmillian.

Heaton, R., Chelune, G., Talley, J., Kay, G., & Curtis, C. (1993). *Wisconsin Card Sorting Test*. Odessa, FL: Psychological Assessment Resources.

Hirsh, M. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, *2*, 331–349.

Hopfield, J., & Tank, W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, *52*, 141–152.

Kaplan, G., Genç, I., & Güzeliş, C. (2001). A simple neural network system for Wisconsin Card Sorting Test. Proceeding of the EMBC 2001.

Kimberg, D., & Farah, M. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*, 411–428.

Knight, R., & Grabowecky, M. (2000). Prefrontal cortex, time and consciousness. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed). (pp. 1319–1339). Cambridge, MA: The MIT Press.

Koren, D., Seidman, L., Harrison, R., Lyons, M., Kremen, W., Caplan, B., et al. (1998). Factor structure of the Wisconsin Card Sorting Test: Dimensions of deficit in schizophrenia. *Neuropsychology*, *12*, 289–302.

Leven, S., & Levine, D. (1987). Effects of reinforcement on knowledge retrieval and evaluation. IEEE first ICNN (Vol. 2) pp. 269–279.

Levine, D., & Prueitt, P. (1989). Modeling some effects of frontal lobe damage—novelty and perseveration. *Neural Networks*, *2*, 103–116.

Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human anatomy and the frontal lobes. Part I: Imitation and utilization behavior. *Annals of Neurology*, *19*, 326–334.

Lombardi, W., Andreason, P., Sirocco, K., Rio, D., Gross, R. E., Umhau, J. C., et al. (1999). Wisconsin Card Sorting Test: Performance following head injury: Dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology*, *21*, 2–16.

Mega, M., & Cummings, J. (1994). Frontal-subcortical circuits and neuropsychiatric disorders. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *6*, 358–370.

Mehrotra, K., Mohan, K., & Ranka, S. (1997). *Elements of artificial neural networks*. Massachusetts: The MIT Press.

Mesulam, M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, *10*, 309–325.

Mesulam, M. (1986). Frontal cortex and behavior. *Annnals of Neurology*, *19*, 320–325.

Mesulam, M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory [review]. *Annals of Neurology*, *28*, 597–613.

Mesulam, M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.

Mesulam, M. (2000a). Attentional networks, confusional states and neglect syndromes. In M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 174–256). New York: Oxford University Press.

Mesulam, M. (2000b). Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes the limbic system and hemispheric specializations. In M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 1–120). Oxford: Oxford University Press.

Miller, E., & Asaad, W. (2002). The prefrontal cortex: Conjunction and cognition. In F. Boller, & J. Graffman (Eds.), *Handbook of neuropsychology* (2nd ed.) (pp. 29–54). Amsterdam: Elsevier.

Miller, E., & Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Milner, B. (1963). Effects of brain lesions on card sorting. *Archives of Neurology*, *9*, 90–100.

Monchi, O., Taylor, J., & Dagher, A. (2000). A neural model of working memory processes in normal subjects, Parkinson's disease and schizophrenia for fMRI design and predictions. *Neural Networks*, *13*, 953–973.

Morecraft, R., Geula, C., & Mesulam, M. (1993). Architecture of connectivity within a cingulo-fronto-parietal neurocognitive network for directed attention. *Archives of Neurology*, *50*, 279–284.

Nelson, H. (1976). A modified card sorting test sensitive of frontal lobe defects. *Cortex*, *12*, 313–324.

O'Reilly, R., Noelle, D., Braver, T., & Cohen, J. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral Cortex*, *12*, 246–257.

Polk, T., Simen, P., Lewis, R., & Freedman, E. (2002). Computational approach to control in complex cognition. *Cognitive Brain Research*, 71–832.

Posner, M. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences USA*, *91*, 7398–7403.

Raichle, M. (2000). The neural correlates of consciousness: An analysis of cognitive skill learning. In M. Gazzaniga (Ed.), *The new cognitive neurosciences*2nd ed. (pp. 1305–1318). Cambridge, MA: The MIT Press.

Raichle, M., Fiez, J., Videen, T., MacLeod, A., Pardo, J., Fox, P. T., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, *4*, 8–26.

Rugg, M. (1997). *Cognitive neuroscience*. East Sussex: Psychology Press.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shepard, D., Bradshaw, J., Purcell, R., & Pantelis, C. (1999). Tourette's and co-morbid syndromes: Obsessive–compulsive and attention deficit hyper-activity disorder. *Clinical Psychology Review*, *19*, 531–552.

Shulman, G., Fiez, J., Corbetta, M., Buckner, R., Miezin, F., Raichle, M., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience*, *9*, 648–663.

Sohn, A., & Gaudiot, J. (1992). Learning legal moves in planning problems: A connectionist approach to examining legal movies in the Tower-of-Hanoi. *Engineering Application of Artificial Intelligence*, *5*, 239–245.

Stedman, T., & Clair, A. (1998). Neuropsychological, neurological and symptom correlates of impaired olfactory identification in schizophrenia. *Schizophrenia Research*, *32*, 23–30.

Stillings, N., Weisler, S., Chase, C., Feinstein, M., Garfield, J., & Rissland, E. (1995). *Cognitive science* (2nd ed.). Massachusetts: The MIT Press.

Stuss, D., & Benson, D. (1986). *The frontal lobes*. New York: Raven Press.

Tekin, S., & Cummings, J. (2002). Frontal-subcortical, neuronal circuits and clinical neuropsychiatry. *Journal of Psychosomatic Research*, *53*, 647–654.

Wolford, G., Miller, M., & Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. *The Journal of Neuroscience*, *20*, 1–4.

Zurada, J. (1992). *Introduction to artificial neural systems*. St. Paul, MN: West Publishing Company.