The neurodynamics underlying attentional control in set shifting tasks

Anja Stemme · Gustavo Deco · Astrid Busch

Received: 3 November 2006 / Accepted: 24 March 2007 / Published online: 5 June 2007
© Springer Science+Business Media B.V. 2007

Abstract In this work we address key phenomena observed with classical set shifting tasks as the “Wisconsin Card Sorting Test” or the “Stroop” task: Different types of errors and increased response times reflecting decreased attention. A component of major importance in these tasks is referred to as the “attentional control” thought to be implemented by the prefrontal cortex which acts primarily by an amplification of task relevant information. This mode of operation is illustrated by a neurodynamical model developed for a new kind of set shifting experiment: The Wisconsin-Delayed-Match-to-Sample task combines uninstructed shifts as investigated in Wisconsin-like tasks with a Delayed-Match-to-Sample paradigm. These newly developed WDMS experiments in conjunction with the neurodynamical simulations are able to explain the reason for decreased attention in set shifting experiments as well the different consequences of decreased attention in tasks requiring bivalent yes/no responses compared to tasks requiring multivalent responses.

Keywords Wisconsin Card Sorting Test · Stroop task · Set shifting task · Stimulus congruency · PFC · Neurodynamical model · Attentional control

Electronic supplementary material The online version of this article (doi: [10.1007/s11571-007-9019-8](https://doi.org/10.1007/s11571-007-9019-8)) contains supplementary material, which is available to authorized users.

A. Stemme (✉) · A. Busch
Department Psychologie, LMU Munich, Leopoldstr. 13, Munich
80802, Germany
e-mail: stemme@psy.uni-muenchen.de; anja.stemme@gmx.de

G. Deco
Computational Neuroscience, Universitat Pompeu Fabra,
Passeig de Circumvallació 8, Barcelona 08003, Spain
e-mail: Gustavo.Deco@upf.edu

Introduction

Traditionally, in set shifting tasks like the “Wisconsin Card Sorting Test” (WCST, Grant and Berg 1948) or the Stroop task (Stroop 1935) the ability of the participants to switch attention from one aspect of an object to another is investigated. In these tasks subjects are required to attend to a selected property of a presented visual stimulus and provide an according feature specific response. The different stimulus properties might interfere each other as in the Stroop task: written color(ed) words serve as stimulus displays and the participants are instructed to switch between the response rules “color naming” and “word reading”. Congruent and incongruent stimulus displays are differentiated: A congruent stimulus display is constituted for example by the color word “RED” written with red ink whereas an incongruent stimulus display might be the color word “RED” written with green ink. The valid response rule is usually indicated by an explicit task cue or a pre-defined task order. In the WCST participants are required to sort cards according to one of three possible sorting criteria: color, form or number. The stimulus properties do not interfere each other as in the Stroop task but the valid sorting rule is changed without explicit notice and the participants are required to detect the new valid rule by a trial and error procedure.

Various phenomena have been observed with set shifting tasks: increased response times following set shifts as well as following the presentation of congruent visual stimuli. Also, for incongruent stimulus displays compared to congruent ones response times are increased (see Monsell 2003, for an overview). Further more, increased error rates were detected following an uninstructed change of the valid rule especially for patients suffering from lesions of the pre-frontal cortex (e.g. Milner 1963; Barceló and Knight 2002).

A key component with respect to the phenomena observed with set shifting tasks is the “attentional control” thought to be implemented by the prefrontal cortex. This attentional control is responsible for the preference of task relevant stimuli (or stimulus features) against irrelevant ones and the switch of this preference if requested by an implicit or explicit demand to switch the valid response rule. Egner and Hirsch (2005) conducted neuroimaging studies which revealed that this attentional control acts more likely by an amplification of task-relevant information rather than an inhibition of distracting stimuli (or stimulus properties).

In Stemme et al. (2007) we addressed the complex of set shifting tasks by the presentation of a new type of set shifting experiment, called “Wisconsin-Delayed-Match-to-Sample” task (WDMS task), and according neurodynamical simulations. The WDMS task allows the investigation of different aspects of set shifting tasks though using a rather simple setup to enable realistic neurodynamical simulations. The experimental setup combines a visual “Delayed-Match-to-Sample” task with a Wisconsin-like paradigm: Two visual stimuli, consisting of simple colored shapes, are presented to the subjects, separated by a delay. Following the presentation of the second stimulus the participants are required to state whether the stimuli matched with respect to a given criteria or did not match. Two different possible matching criteria were chosen and the relevant one was changed at arbitrary intervals without explicit notice (uninstructed Wisconsin-like set shift). This setup allowed to investigate effects of stimulus congruency (comparable to Stroop tasks) as well as uninstructed set shifts (investigated in WCST tasks). The simulation results matched the experimental performance measures for the WDMS task in terms of subject response times as well as with respect to different error types. In considering not only experimental average results but as well individual variations we obtained a rather detailed description of human behavior compared to previous studies where only average results were considered, normally without standard deviations (see e.g. Stuss et al. 2000, 2001; Gilbert and Shallice 2002; Monsell 2003).

The neurodynamical model developed for the WDMS simulations suggested an asymmetric set of synaptic connections to account for a memory based switching process. Stronger feedforward connections are responsible for the maintenance of a valid response rule and stronger feedback connections are responsible for the selection of an alternate rule and hence for the shift of attention. An important property of this model is that the selected weight set remains fixed during the simulations. Thus, the necessary set shift during the simulations does not require a weight modification process or any other artificial additions (explicit excitation of a new rule, delay of the feedback

information, etc., compare as well Rougier and O'Reilly 2002) opposed to existing more abstract neuronal models (Dehaene and Changeux 1991; Braver et al. 1999; O'Reilly et al. 2002; Rougier et al. 2005). If, for example, weight modifications are assumed to be necessary for every set shift this would imply that subjects actually *learn* every application of an abstract rule again and again. Such a learning and relearning of familiar tasks seem to be at least highly questionable and therefore we consider a fixed set of weights during the simulations as an important model advantage.

The *degree* of attention is reflected within the model by arbitrarily varying external fluctuations which are responsible for different types of errors. These errors represent a system inherent model feature and are comparable to those errors committed by healthy subjects during the experiments. The response times generated by the model were dependent on the stimulus congruency comparable to the effects observed in Stroop-like tasks (see e.g. Monsell 2003; Egner and Hirsch 2005).

In this work we address the question of decreased attention *following* the presentation of congruent stimuli. In classical set shifting tasks decreased attention following trials with congruent stimuli is reflected in increased response times (e.g. Monsell 2003; Egner and Hirsch 2005). However, for the WDMS task we detected slightly increased error rates rather than increased response times following congruent trials. The neurodynamical model of the WDMS task is able to explain the neuronal reasons for the decrease of attention and, further more, allows to hypothesize and explain why decreased attention is reflected in increased response times for Stroop-like tasks but in increased error rates in tasks requiring bivalent yes/no responses.

A range of studies already proved successfully that biophysically detailed neurodynamical models are a valuable tool to gain an understanding about the neuronal basics of selective working memory and attention (Corchis and Deco 2002; Deco and Rolls 2003, 2005; Deco et al. 2004; Almeida et al. 2004). Further more, in a previous study (Stemme et al. 2005) we demonstrated the calculation of event related fMRI signals for set shifting tasks. The entire complex of set shifting tasks and biophysical models is discussed in detail as well in Stemme (2007).

Materials and methods

Wisconsin-DMS experiments

In Fig. 1 the experimental setup is depicted (Compare as well Stemme et al. (2007); Stemme (2007)). A sample display was shown for 500 ms, followed by a fixation delay

Fig. 1 Setup of the “Wisconsin Delayed-Matching-to-Sample” experiments (top) and example trial sequence (bottom). The trial sequence shows a rule change in the second trial; the object rule was valid in the first trial. Starting with the second trial the new valid rule is the space rule

of 1,000 ms, followed by a test display which was presented until the subjects responded by key press (“y”—yes or “n”—no; yes—sample and test display matched with respect to the valid rule, no—sample and test display did not match according to the currently valid rule). Afterwards a feedback message informed the subjects whether their response was “correct” or “wrong”. The feedback message was presented for 1,500, 1,000 or 500 ms (W-DMS I, II, III). The subjects had to discriminate between two different possible rules: Same position on the screen (called further on “space rule”) or same object presented in sample and test display (with respect to all feature dimensions; called “object rule” in the reminder of the text). Sixty-four colored rectangles on 64 different possible positions on the screen served as stimulus material. After an arbitrary number of correct trials (3, 5, 7, 9 or 11 not necessarily consecutive correct trials) the valid rule was changed without notice (Wisconsin-like paradigm). In this case the subjects received the feedback message “wrong” although they responded correctly according to the valid rule in the previous trial (see Fig. 1, bottom). The experiments were conducted with 40 healthy participants, one experiment comprised about 850 single trials.

Four different match conditions of the visual stimuli were differentiated (see as well Fig. 1, top right):

both: The stimulus presented in the sample display and the one presented in the test display are identical with respect to both rules.

match: The stimulus presented in the sample display matches the test display stimulus *only* with respect to the currently valid rule.

nonmatch: The stimulus presented in the sample display does *not* match the test display stimulus with respect to the currently valid rule.

none: The stimulus presented in the sample display and the one presented in the test display are different with respect to both dimensions.

Trials with “both” and “none” conditions are also referred to as trials with *congruent* stimulus conditions (short “congruent trials”) whereas trials with “match” and “nonmatch” conditions are summarized as trials with *incongruent* stimulus conditions (short “incongruent trials”). In congruent trials both feature dimensions require the same response i.e. “yes” for trials with both conditions and “no” for trials with none conditions; in incongruent trials the two feature dimensions require different responses (compare as well Fig. 1). This differentiation is very similar to the stimulus differentiation tested in Stroop-like tasks (see e.g. Monsell 2003) though the WDMS design provides further differentiation possibilities (two different congruent conditions and two different incongruent condition) whereas Stroop tasks only permit the comparison of “both” conditions with “match” conditions however with a stronger interference of the possible rules.

The neurodynamical model

The neurodynamical model (see as well Stemme et al. 2005, 2007; Stemme (2007)) is based on the framework first introduced by Brunel and wang (2001). The model represents a selected cortical area and consists of two different types of neurons, 80% excitatory pyramidal cells and

20% inhibitory interneurons, consistent with neurophysiological findings (Abeles 1991). The neurons are grouped into two appropriate types of pools. Each neuron is modelled as an “Integrate-and-Fire” neuron taking into account three different synaptic connection types: two excitatory—AMPA and NMDA connections—and one inhibitory—GABA. The different synaptic connection types are represented and computed using equivalent electrical circuits, consisting basically of a resistance parallel to a conductance with type specific parameter values for conductivity and resistance following experimentally determined values (Brunel and Wang 2001).

Every neuron receives a certain background input from neurons outside the network modelled. The cerebral cortex is highly connected and thus the simulation of a “closed” cortical area would be unrealistic. For the approximation of the background input, it is taken into account that neurons always show a certain level of activity, i.e. a spiking rate of approximately 3 Hz for pyramidal cells and 9 Hz for interneurons, which is called the “spontaneous rate”, (see e.g. Wilson et al. 1994). Accordingly, the external background input is modeled as an AMPA-mediated Poisson train of spikes arriving from $N_{ext} = 800$ neurons with a rate of 3 Hz. Thus, the total background noise each modeled neuron receives comes out to $v_{ext} = 800 * 3 \text{ Hz} = 2.4 \text{ kHz}$.

The neurodynamical model (compare Fig. 2) considers 1,600 excitatory neurons and 400 inhibitory neurons. These neurons are grouped into pools which are responsive for a certain stimulus feature (feature pools: O1, O2, S1, S2) or represent an abstract rule (rule pools: RO, RS). Each of these “selective” pools incorporates 100 excitatory neurons. The remaining excitatory neurons are organized into the pool of non-selective neurons (not all neurons within a given cortical area respond to a specific task, see e.g. White and Wise 1999). This pool also introduces some noise and supports the generation of the almost Poisson-like firing patterns of the neurons in the simulation which is a property of many neurons observed in the cortex (Brunel and Wang 2001). The inhibitory neurons are grouped to form one inhibitory pool which implements a global competition between all neurons within a given cortical area again consistently with experimental findings. The network of neurons is fully connected with different connection strengths.

The identification of the various pools was supported by single cell recordings with behaving monkeys. These recordings demonstrated that single neurons show rule specific (e.g. Wallis et al. 2001; White and Wise 1999) as well as object specific (e.g. Rainer and Miller 2002) activity in a range of behavioral tests. These results

Fig. 2 The neurodynamical model for the simulations covering 1,600 excitatory and 400 inhibitory neurons. I = Pool of inhibitory neurons. N = pool of non-selective neurons. RO = neuronal pool for the object rule, RS = pool for the space rule. O1/O2 = object pools, representing two different objects; S1/S2 = space pools, representing two different positions on the screen. Each selective pool comprises

100 neurons. All neurons receive an AMPA-mediated external input of 2.4 kHz. Additional external AMPA-mediated input is provided to the pools in order to simulate the desired task. The asymmetry of the weight set is underlined by the usage of thicker arrows for stronger connections between rule and feature pools (refer to the text)

substantiate the assumption that groups of neurons (i.e. the pools) code for specific stimulus features as well as for the more abstract rules in the tasks we aim to simulate. Hence, the model comprises two pools serving as “rule pools”, representing two different, possible active rules and four stimulus specific pools, representing two times two different stimulus properties: Two different objects (“Object Pools”) which may be presented at two different locations on the screen (“Space Pools”).

An object at a certain location is presented to the model (e.g. object number one at position number two) by adding an extra Poisson input to the specific pools (object pool No. 1 and space pool No. 2). For this purpose the external AMPA-mediated input to the neurons within the specific pool is increased to $v_{ext} + \lambda_{stimulus}$. Compared to the background noise, $v_{ext} = 2.4$ kHz, the stimulus specific input is rather low: $\lambda_{stimulus} = 0.15$ kHz.

To raise and hold competition the rule pools receive continuously a low attentional biasing input, $\lambda_{bias} = 0.07$ kHz. One of the rule pools wins this competition and enters into a state of persistent activity with a spiking frequency of about 20–30 Hz. This active, spiking rule pool represents the rule “selected” by the model to be currently valid. At the end of a simulated trial we introduce an unspecific extra external input to the model representing the feedback the model would receive to the previously given answer. The feedback input is provided simultaneously to both of the rule pools, thus v_{ext} is increased by λ_{bias} and $\lambda_{feedback}$. In case of a correct answer we refer to the feedback input as “positive feedback” and “negative feedback” in case of an incorrect answer. However, the

feedback input itself is in both cases an external, unspecific AMPA mediated input to both of the rule pools, differing just in the amount: $\lambda_{posFB} = 0.05$ kHz and $\lambda_{negFB} = 0.45/0.55$ kHz. The positive feedback actually acts as a “strengthenener” of a currently active, i.e. spiking, rule pool whereas the higher negative feedback destabilizes the rule pools and hence erases any previously persistent activity of any of the two pools. Boundaries of these values for $\lambda_{feedback}$ were evaluated in exemplary simulations.

The rule switch is implemented as a memory based switching process (compare as well Fig. 3) and accomplished by the model in a very natural way without artificial additions by the usage of an asymmetric set of weights. In the “feedforward” direction, every rule pool supports “its” feature pools via a comparatively strong weight $w_{rf} \gg w_{non}$ (see Fig. 2). Hence, the object rule pool (RO) provides a higher amount of input to the object feature pools (O1 and O2) than to the space feature pools (S1 and S2). Contrary, in the opposite, “feedback” direction the *object* feature pools (O1 and O2) provide a higher amount of input to the *space* rule pool (RS) than to the object rule pool (RO); thus $w_{nonr} \gg w_{fr}$. This asymmetric connection configuration enabled the model to switch the rules or shift the set, respectively.

The success of a set shift depends on the degree of *activity difference* between the feature pool representing the relevant stimulus feature compared to the one representing the irrelevant one. An example of the spiking dynamics during the simulations is provided in Fig. 1, supplemental material. Fluctuations in the external AMPA input (v_{ext}) might overlay the activity difference preventing

Fig. 3 Illustration of the switching process. At the beginning of the trial the space rule (RS) is assumed to be active and a stimulus number 1 is presented by the external excitation of the corresponding stimulus pools (O1 and S1). During the delay period the relevant stimulus dimension is memorized and thus pool S1 shows a higher level of activity compared to pool O1 (second picture). Afterwards the second stimulus is presented to the model (O2 and S1). Due to a wrong model response the negative feedback is provided to both of the rule pools. The fifth picture illustrates the destabilization of the rule pools. Hence, both of the pools show only a low level of activity

where now the rule pool representing the irrelevant rule (RO in this case) has an advantage: This rule pool receives a higher input from the stimulus pools than the previously active rule pool. Thus in the next trial we see the opposite rule pool active (sixth picture, pool RO). However, this switching procedure does not necessarily succeed: Fluctuations in the external AMPA input influence the spiking activity of the pools and prevent on the one hand the proper memorization of the stimulus features (leading to wrong model responses) and prevent on the other the successful completion of the rule changes

the successful completion of a rule change. As a consequence, the model elicits as well perseverative errors during the simulations.

The actual model response is determined algorithmically based on the spiking activity of all four feature pools (summed spiking rate, ssr):

- The model response is considered to be “yes” if the ssr crossed a threshold T_{yes} x number of times;
- accordingly the model response is considered to be “no” if the ssr stays y number of times *below* a threshold T_{no} .

This way of response determination accounts for the effect that the feature pools together (respectively the summed spiking rate of all feature pools) show a comparatively high spiking rate for trials requiring a “yes” response and a lower one for “no” responses. The according spiking dynamics are illustrated as well in Fig. 1, supplemental material. The parameter values for x , y , T_{yes} and T_{no} were determined in evaluative simulations aiming in the generation of a maximal amount of correct model responses using exemplary trial sequences (compare as well Stemme et al. (2007)).

Further on, an optional minimal spiking rate of the feature pools was considered in order to generate a neuronal response. This threshold, T_{min} , accounted especially for the “No”-responses of the model as in this case the spiking rate is required to stay *below* a certain value. Hence, the threshold T_{min} represents the assumption that a certain spiking level might be necessary for the neuronal spiking development to become “conscious”.

In summary, a response of the model is considered to be “yes” (R_{yes}):

$$R_{yes} = (ssr >_x T_{yes}) \wedge (ssr >_1 T_{min}) \quad (1)$$

and accordingly considered to be “no” (R_{no}):

$$R_{no} = (ssr <_y T_{no}) \wedge (ssr >_1 T_{min}) \quad (2)$$

The response time of the model equals the time when $R_{yes/no}$ was finally decided: $t(R_{yes/no})$.

Several options are possible for the neurodynamical WDMS-model (see Table 1, compare as well Stemme et al. (2007)). These options include two alternative sets of asymmetric weights to enable the set shift as well as two alternate possibilities to implement the feedback provision. Furthermore, the fluctuations in the external AMPA input were varied. As the fluctuations are generated with the support of a random number generator these may be varied by the usage of different random seeds or “controlled” by using the same fixed random sequence for every trial. Finally different threshold sets are possible for the response

determination. Hence, altogether 15 simulations, comprising 300 trials each, with different model options served as a principal data base for the WDMS task. For all simulations we used a feedback time of 1,000 ms as the experimental results revealed no effect of the feedback duration on response times or error rates.

Results

Response times and error rates are depicted in Fig. 4. Simulations 1, 7 and 13 were chosen exemplarily to illustrate the principal match of the simulation results with the average experimental results.

Thus, we see increased response times for incongruent compared to congruent trials comparable to the effects commonly observed with Stroop tasks (e.g. Gilbert and Shallice 2002; Monsell 2003). However, this item appears to be true only for congruent trials with “both” conditions but not for the trials with “none” conditions. This circumstance was analyzed in detail in Stemme et al. (2007) by the consideration of individual experimental results and further explained and substantiated by the neurodynamical simulations. In this respect a striking effect turned out to be that the usage of the additional minimal threshold inverted the relationship between the response times for “none” and “nonmatch” trials (see Fig. 4, simulation 7).

In Stemme et al. (2007) we demonstrated more over that simulations using the described model are able to cover individual experimental results with respect to average values and standard deviations. The investigations allowed the identification of characteristic response profiles for the participants of the WDMS experiments. These profiles were reproduced and explained by the neurodynamical simulations. Thus, we were finally able to conclude that all investigated model options as listed in Table 1 are plausible and represent individual *neuronal variation details* around the central concept of set shifting.

In Fig. 4, middle diagrams the response times relative to a set shift averaged across all trials following a rule change are outlined. Compared to the increased response times in common set shifting tasks following a set shift (about ~40% or 200 ms to a baseline of 500 ms, compare e.g. Monsell 2003) the increase was moderately low for the experiments as well as for the simulations. Response times were increased by only 12–14% for switch trials compared to non-switch trials. Further on, in the remaining repetition trials (RC X) response times increased again for the simulations as well as for WDMS I and III.

For the analysis of error rates we differentiated the following error types (compare Fig. 4, right diagrams): The first errors of an experiment or a trial block within an

Table 1 Overview of task model options used for the experimental simulations

No	w_{mon}	FBT (ms)	λ_{negFB} (Hz)	Random seed	Start rule	$T_{yes/no}$	x	y
1	1.2	200	450	0	Object	55	3	12
2	1.2	200	450	0	Space	55	3	12
3	1.2	200	450	1	Object	50	5	11
4	1.2	200	450	1	Space	55	3	12
5	1.2	200	450	111 (*)	Space	50	3	11
6	1.3	200	450	0	Space	55	3	12
7	1.3	200	450	0	Space	50	4	10
8	1.3	200	450	1	Object	55	3	12
9	1.3	200	450	1	Space	55	3	12
10	1.3	200	450	111(*)	Space	50	3	11
11	1.3	150	550	0	Space	55	3	12
12	1.3	150	550	0	Space	50	3	9
13	1.3	150	550	1	Object	55	3	12
14	1.3	150	550	1	Space	55	3	12
15	1.3	150	550	17 (*)	Space	55	4	13

For simulations 1–5 w_{mon} was set to 1.2, for the remaining 10 simulations the higher value (1.3) was used to include unmotivated errors due to arbitrary rule changes. The higher value of this weight enables arbitrary rule changes. This means that a previously inactive rule pools becomes active without any prior destabilization of the persistent spiking activity. The task model using $w_{mon} = 1.2$ is more robust in the maintenance phases. Simulations 1–10 used a longer actual feedback time and accordingly a slightly lower value for the actual feedback compared to the remaining five simulations. Simulations 1, 5 and 10 were conducted with controlled fluctuations which means that the same random number sequence for the generation of the external AMPA input was used in *every* trial. For the remaining simulations the external fluctuations were varied by different random seeds or different valid starting rule. The minimal threshold for all simulations was set to 0. For simulation 7 (compare Fig. 4) T_{min} was set to 55 Hz. The remaining model parameters are provided as supplemental material

Fig. 4 Results in terms of response times and error rates for the W-DMS experiments (bottom diagrams) and the simulations (top diagrams). Average response times for the different match conditions (left diagrams) and relative to the rule change (middle diagrams) as well as error rates and types (right diagrams). For the W-DMS experiments three variants with different feedback times were

considered: 1,500 ms (W-DMS I), 1,000 ms (W-DMS II) and 500 ms (W-DMS III). Note: For all three simulations an average feedback time of 1,000 ms was assumed. RC1: First trial after the rule change, RC2: second trial after the rule change, RC X the remaining trials. RCF: Errors following a rule change, UE: unmotivated error, UEF: error following an unmotivated error, AQ: rule acquisition error

experiment where subjects try to find out the first valid rule were called “rule acquisition errors” (AQ). These errors were considered to be of minor relevance as the simulations did not consist of multiple blocks. The necessary errors subjects commit in the rule change trial when the valid rule

is changed without notice (see Fig. 1, bottom, trial 2) were not considered as errors in its original sense and thus excluded from the analysis. Errors in the context of a rule change were termed “RCF” (Rule Change Follow up). These are errors occurring after the valid rule was changed

and before the new valid rule is considered to be definitely established by the subjects. The establishment of a new rule is assumed to have happened after *three* consecutive correct answers¹ (start of “maintenance phase” of a valid rule). RCFs might be considered as well as “perseverative errors”. Further more, unmotivated errors (UE) are errors conducted after a new rule is considered to be established. These errors were considered to be related to the attention of the participants and called “unmotivated” because the exact reason for these errors was so far undetermined. They are called also “random”, “occasional” or “non-perseverative” errors in the literature (e.g. Barceló and Knight 2002; Rougier and O’Reilly 2002; Rougier et al. 2005). Further on we differentiated errors in the context of an unmotivated error (UEF). These are errors following an unmotivated error. Similar to RCFs, errors were considered as UEFs before the subject again responded correctly in *three* consecutive trials after an UE.

The overall amount of errors was moderately low, about 10–13% for the selected simulations and 6–9% in the experiments. About 40–50% of these errors occurred in conjunction with a rule change (RCF) and ~30–40% of the errors were classified as “unmotivated errors”. The portion of errors categorized as “UEF” was comparatively low (about 10%) as was the portion of acquisition errors (AQ). These results indicated that perseverative errors (RCF) form the greatest part of all of errors for the simulations as well as for the experiments.

Further more, the analyzation of the spiking dynamics of individual simulations suggested that a true perseverative error would more likely occur *immediately* after a set shift rather than within three consecutive trials. This means that a new rule is quite reliably established already with the first correct answer following a rule change. Hence, the neurodynamical WDMS model suggested to differentiate between sequences of errors and single errors rather than the error types named above. Such error sequences might occur within the maintenance phase of an active rule *as well as* following a set shift. Both of these occurrences would indicate perseveration tendencies. On contrary, attentional errors would more likely occur as individual single errors.

Thus, for subsequent error analysis single errors (SE) and error sequences (ES) were distinguished and analyzed for the simulations as well as the experiments. Again similar error profiles showed up for the simulations as for the experiments with single errors now constituting the biggest part of all errors committed for the majority of experimental participants and the according simulations.

However, one striking effect for the WDMS experiments as well as the simulations turned out to be the rather

low increase in response times following the set shift and, contrary to traditional set shifting experiments, the response times were *not* increased following congruent trials. Such an increase is thought to indicate decreased attention as confirmed in neuroimaging studies (see Egner and Hirsch 2005). This is an important aspect as the overall operation of the model is well compliant with the findings described in this study.

We hypothesized that the reason for this missing effect is laid in the design of the experiment with bivalent yes/no responses. From a neuronal point of view the development of the summed spiking rate (*ssr*) of all feature pools is responsible for the generation of an actual response for the simulations as for the subjects. If a single threshold ($T_{yes} = T_{no}$) is assumed to discriminate between “yes” and “no” responses the neurodynamics leave only comparatively little room for increased response times as this single threshold is either passed or failed and each of these events contributes to the response generation. Thus, the usage of a single threshold $T_{yes} = T_{no}$ (or accordingly bivalent rather than multivalent responses in the experiments) would predict that decreased attention is reflected in an increase of error rates rather than an increase of response times.

This means, we should see increased error rates rather than increased response times following congruent trials. To investigate this item we analyzed again the error rate distribution. Taking into account the reduced amount of congruent trials compared to incongruent trials within the experiments, we compared the rate of congruent trials (supposed to be about 25% of all trials) with the portion of unmotivated and single errors occurring *after* a congruent trial in the simulations. If there is no influence of the congruent trials on the error rates in the following trial it is hypothesized that as well only approximately 25% of the unmotivated or single errors occur after a congruent trial.

Figure 5 shows the result for all 15 individual simulations. Almost all simulations suggest that the portion of unmotivated errors occurring after congruent trials is higher than the overall portion of congruent trials. This difference even increases when considering single errors and thus further substantiates the suggested differentiation of single errors and error sequences.

The analysis of the experimental data delivered similar results (Fig. 6, right diagram). Hence, the portion of unmotivated errors committed after congruent trials was significantly larger than the portion of congruent trials itself. The same applies for the portion of single errors. Thus, a disproportional amount of attentional errors is committed by the participants following congruent trials. Although of course overall only a minority of congruent trials led to an error in a subsequent trial the obtained results indicate at least that congruent trials in the WDMS task have a

¹ Similar to Konishi et al. (1999) and Nakahara et al. (2002) who considered a new rule to be established after three correct trials.

Fig. 5 Error result for all 15 individual simulations. Fourteen of the 15 simulations (comprising 300 trials each) suggest that the portion of unmotivated errors (UE) occurring after congruent trials is higher than the overall portion of congruent trials. For 12 simulations this difference even increases when considering single errors (SE) and

thus further substantiates the suggested differentiation of single errors and error sequences. Simulation number 5 used controlled external fluctuations and thus produced only one error at all which was a perseverative error. AV: average across all simulations

Fig. 6 Average error analysis for the simulations (left diagram) compared to the WDMS experiments (WDSM I, II and III, right diagram). For the experiments, two-tailed *t*-tests revealed a significant difference (−6.9%) for the amount of congruent trials compared to the amount of unmotivated errors after congruent trials ($t(39) = -3.7$, $P = 0.0007$) and a significant difference of −9.99% for the amount of

congruent trials compared to single errors after congruent trials ($t(39) = -7.4$, $P = 6.213e-09$). The difference between the portion of unmotivated and single errors after congruent trials was only close to significant ($P = 0.05846$). cT—percentage of congruent trials; UE—percentage of unmotivated errors following a congruent trial; SE—percentage of single errors following a congruent trial

tendency to weaken the attention of the participants comparable to the effects observed with classical set shifting tasks.

Discussion

In this work we described the neurodynamical simulation of an experiment (Wisconsin-Delayed-Match-to-Sample task, WDMS task) that combines several aspects of set shifting tasks. Two visual stimuli were presented to the subjects, separated by a delay. Following the presentation of the second stimulus the participants were required to state whether the stimuli matched with respect to a given criteria or did not match. Two different possible matching criteria (relevant response rules) were selected and the relevant one was changed at arbitrary intervals without explicit notice (uninstructed Wisconsin-like set shift). In particular, the WDMS allowed to investigate effects of stimulus congruency (comparable to Stroop tasks) as well as set shifting (investigated in WCST tasks). The simulations were able to cover experimental results with a high degree of detail in terms of response time and error rate distributions.

Further more, studies using a more classical Stroop-like set shifting reported increased response times following the presentation of congruent stimuli. Egner and Hirsch (2005) were able to demonstrate that this effect is related to the operation of attentional control within prefrontal cortex, i.e. the amplification of task relevant stimuli. Following congruent trials task relevant stimuli are amplified to a lesser extend than following incongruent trials which actually represents the decreased attention following congruent trials. In the WDMS we did not detect increased response times following congruent trials but slightly increased error rates.

We were able to demonstrate that the reason for this effect is on the hand compatible with the suggested operation of the prefrontal cortex and on the other hand motivated by the design of the experiment with bivalent yes/no responses. The WDMS experiments as well as the simulations suggested that, if experimental participants are required to provide bivalent responses, decreased attention is reflected in increased error rates rather than increased response times. From a neuronal point of view the reason is relatively simple: If a stimulus with various features (or a compound stimulus) is presented to the subjects and they are required to simply *select* the rele-

vant feature, response times increase upon decreased attention until the neuronal activity of the corresponding feature pool reached a certain activity level to generate a response. On contrary, if participants are required to decide (yes/no) whether two stimuli match or do not match the feature pool activity might reach a spiking pattern indicating a wrong response earlier rather than generating a late correct response.

From the neurodynamical point of view the decreased attention might be explained as follows: During congruent trials both rule pools receive a comparatively high (for trials with “both” conditions) or low input (“none” conditions) from the feature pools. Both of these “extreme” input conditions weaken the active, spiking rule pool (i.e. weaken the selective “attention”) in providing too much input as well to the opposite rule pool (asymmetric set of weights) or an overall comparatively small amount of input to the rule pools. As a consequence the spiking activity of the relevant rule (pool) decreases and thus the attention payed to the relevant stimulus dimension which in turn increases the *probability* of an error in the following trial. Similar, in experiments requiring multivalent responses the decreased attention and accordingly decreased spiking activity of the rule pool would lead to a lower spiking activity of the feature pool representing the correct response. Hence, in these experiments the response times increase until the spiking level of the feature pool finally reached a given response threshold.

The results obtained in this work proved that the choice to use a biophysical detailed model comprising “Integrate-and-Fire” neurons was very appropriate. However, future work will have to show whether corresponding expressive results are possible as well with simpler, rate based models (e.g. Wong and Wang 2006) or whether the usage of an even higher degree of detail has any further advantage (e.g. Hodgkin and Huxley 1952; Meunier and Segev 2002). However, up to now, H-and-H models would be computationally too expensive for a realistic simulation of set shifting tasks.

Acknowledgments This work was supported by the ‘Deutsche Forschungsgemeinschaft’ (DFG), Project Number FOR 480/1-1. Very special thanks are directed to Werner X. Schneider, University of Munich, Department Psychology for valuable discussion and advise.

References

- Abeles A (1991) *Corticonics*. Cambridge University Press, New York
- Almeida R, Deco G, Stetter M (2004) Modular biased-competition and cooperation: a candidate mechanism for selective working memory. *Euro J Neurosci* 20:2789–2803
- Barceló F, Knight RT (2002) Both random and perseverative errors underlie worst deficits in prefrontal patients. *Neuropsychologia* 40:49–356
- Braver TS, Barch DM, Cohen JD (1999) Cognition and control in Schizophrenia: a computational model of dopamine and prefrontal function. *Biol Psychiatry* 46:312–328
- Brunel N, Wang X-J (2001) Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci* 11(1):63–85
- Corchs S, Deco G (2002) Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. *Cereb Cortex* 12(4):339–348
- Deco G, Rolls ET (2003) Attention and working memory: a dynamical model of neural activity in the prefrontal cortex. *Eur J Neurosci* 18:2374–2390
- Deco G, Rolls ET (2005) Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cereb Cortex* 15(1):15–30
- Deco G, Rolls ET, Horwitz B (2004) What and where in visual working memory: a computational neurodynamical perspective for integrating fMRI and single-neuron data. *J Cogn Neurosci* 16:683–701
- Dehaene S, Changeux J (1991) The Wisconsin Card Sorting Test: theoretical analysis and modeling in a neuronal network. *Cereb Cortex* 1(1):62–79
- Egner T, Hirsch J (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat Neurosci* 12:1784–1790
- Gilbert SJ, Shallice T (2002) Task switching: a PDP model. *Cognit Psychol* 44:297–337
- Grant DA, Berg E (1948) A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weight-type card sorting problem. *J Exp Psychol* 38:404–411
- Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol (Lond)* 117:500–544
- Konishi S, Kawazu M, Uchida I, Kikyo H, Asakura I, Miyashita Y (1999) Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. *Cereb Cortex* 9 (7):745–753
- Meunier C, Segev I (2002) Playing the devils advocate: is the Hodgkin Huxley model useful? *Trends Neurosci* 25:558–563
- Milner B (1963) Effects of different brain lesions on card sorting. *Arch Neurol* 9:90–100
- Monsell S (2003) Task switching. *Trends Cogn Sci* 7:134–140
- Nakahara K, Hayashi T, Konishi S, Miyashita Y (2002) Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science* 295(5559):1532–1536
- O’Reilly RC, Noelle DC, Braver TS, Cohen JD (2002) Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cereb Cortex* 12(3):246–257
- Rainer G, Miller EK (2002) Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. *Eur J Neurosci* 15(7):1244–1244
- Rougier NP, Noelle DC, Braver TS, Cohen JD, O’Reilly RC (2005) Prefrontal cortex and flexible cognitive control: rules without symbols. *Proc Natl Acad Sci U S A* 102:7338–7343
- Rougier NP, O’Reilly RC (2002) Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cogn Sci* 26:503–520
- Stemme A (2007) *Neuronal principles underlying cognitive flexibility—a biophysical model for set shifting tasks*. Ph.D. thesis, Technical University Munich. Published by Book On Demand GmbH, Norderstedt, Germany
- Stemme A, Deco G, Busch A, Schneider WX (2005) Neurons and the synaptic basis of the fMRI signal associated with cognitive flexibility. *NeuroImage* 26/2:454–470

- Stemme A., Deco G, Busch A (2007) The neuronal dynamics underlying cognitive flexibility in set shifting tasks. *J Comput Neurosci*. doi: 10.1007/s10827-007-0034-x
- Stroop JR (1935) Studies of interference in serial verbal reactions. *J Exp Psychol* 18:643–662
- Stuss DT, Floden D, Alexander MP, Levine B, Katz D (2001) Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. *Neuropsychologia* 39:771–786
- Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, Murphy KJ, Izukawa D (2000) Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia* 38:388–402
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956
- White IM, Wise SP (1999) Research article: rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126(3): 315–335
- Wilson F, Scalaidhe SP, Goldman-Rakic PS (1994) Functional synergism between putative-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proc Natl Acad Sci U S A* 91:4009–4013
- Wong KF, Wang XJ (2006) A recurrent network mechanism of time integration in perceptual decisions. *J Neurosci* 26(4):1314–1328