Running head: Dynamic Goal States

Dynamic Goal States: Adjusting Cognitive Control without Conflict Monitoring

Stefan Scherbaum, Maja Dshemuchadse, Hannes Ruge, Thomas Goschke

Technische Universität Dresden

Keywords: cognitive control, conflict adaptation, conflict monitoring, anterior cingulate cortex, multistable attractor dynamics, goal representations

Correspondence should be addressed to:

Stefan Scherbaum

Department of Psychology, Technische Universität Dresden

Zellescher Weg 17, 01062 Dresden, Germany

Phone: ++49 351 463 33598

E-mail: Stefan.Scherbaum@psychologie.tu-dresden.de

Abstract

A central topic in the cognitive sciences is how cognitive control is adjusted flexibly to changing environmental demands at different time scales to produce goal-oriented behaviour. According to an influential account, the context-sensitive recruitment of cognitive control is mediated by a specialized conflict monitoring process that registers current conflict and signals the demand for enhanced control in subsequent trials. This view has been immensely successful not least due to supporting evidence from neuroimaging studies suggesting the conflict monitoring function is localized within the anterior cingulate cortex (ACC) which, in turn, signals the demand for enhanced control to the prefrontal cortex (PFC). In this article, we propose an alternative model of the adaptive regulation of cognitive control based on multistable goal attractor network dynamics and adjustments of cognitive control within a conflict trial. Without incorporation of an explicit conflict monitoring module, the model mirrors behavior in conflict tasks accounting for effects of response congruency, sequential conflict adaptation, and proportion of incongruent trials. Importantly, the model also mirrors frequency tagged EEG data indicating continuous conflict adaptation and suggests a reinterpretation of the correlation between ACC and the PFC BOLD data reported in previous imaging studies. Together, our simulation data propose an alternative interpretation of both behavioral data as well as imaging data that have previously been interpreted in favor of a specialized conflict monitoring process in the ACC.

1. Dynamic Goal States: Adjusting Cognitive Control without Conflict Monitoring

While on your way to the kitchen to get a knife, coming across the watering can you start to water your plants, wondering later what you originally intended to do in the kitchen. This everyday example illustrates the importance of cognitive control processes when we pursue goals in environments that contain abundant distracting affordances. Focusing on goals is necessary to perform non-routine tasks successfully, especially in the face of competition from prepotent habitual or stimulus-triggered responses. In experimental psychology, the context-sensitive recruitment of cognitive control processes has often been studied by using conflict tasks.

In such conflict tasks, the basic recruitment of cognitive control processes is measured by the size of the “congruency effect”, that is, the difference in response time (RT) between incongruent (response conflict) and congruent (no response conflict) trials. This congruency effect is supposed to reflect the additional time necessary to resolve increased response conflict in incongruent as compared to congruent trials. For instance, in the Stroop task (Stroop, 1935) participants have to name the ink color of a color word (e.g. ‘red’) while ignoring the word meaning (e.g. ‘blue’). When the color and the word meaning are incongruent, RT is increased compared to trials on which color and word meaning are congruent. Similarly, in the Flanker task (Eriksen & Eriksen, 1974), participants have to respond to a centrally presented target (e.g. letters), but have to ignore surrounding flanker stimuli, which typically produce increased RT when the response indicated by the target and the response indicated by the flanker stimuli are incongruent.

Importantly, for the mentioned conflict tasks it has been found that the congruency effect (i.e., reflecting additional time needed to resolve conflict) can be modulated by additional context variables implicating processes of *control adjustments* working on different time scales.

On a short time scale, conflict in the previous trial modulates the recruitment of control, leading to a reduced congruency effect following incongruent trials (Gratton, Coles, & Donchin, 1992). Hence, this ‘conflict adaptation effect’ could reflect a process that adjusts control from trial to trial. On a longer time scale, the overall frequency of incongruent trials modulates the basic recruitment of control, leading to an increased congruency effect for a lower overall frequency of incongruent trials (Gratton et al., 1992; Tzelgov, Henik, & Berger, 1992). Hence, this ‘congruency proportion effect’, could reflect a process that adjusts control across a series of trials.

2. Conflict Monitoring Theory

A major step towards a mechanistic explanation of the context-sensitive regulation of cognitive control was taken when Botvinick and colleagues (Botvinick, Braver, Barch, Carter, & Cohen, 2001) presented the conflict monitoring theory, which is still one of the most influential attempts to account for flexible adjustments of cognitive control. This theory offered a powerful and parsimonious account of the context-sensitive regulation of cognitive control with its central assumption that a specific neural module, localized within the anterior cingulate cortex (ACC), detects response conflicts which serve as a signal indicating an increased demand for cognitive control to the prefrontal cortex (PFC). As a consequence, conflicts trigger an enhanced mobilization of cognitive control and increased activation of relevant goals represented in the PFC, in order to be better prepared for a subsequent encounter with conflicting sources of information on the next trial of the task. In their model, Botvinick and colleagues (2001) assumed that the co-activation of incompatible responses at the end of a trial is used as the conflict signal for the adjustment of control, which thereby biases processing towards task-relevant information on the next trial. This way, they provided an elegant account of the recruitment of cognitive control. One of the strengths of conflict monitoring theory was its ability to integrate different findings into a common framework. Hence, the theory explained conflict adaptation and congruency proportion effects.

In recent years, this fruitful theory (Botvinick, Cohen, & Carter, 2004) has been supported by imaging studies showing correlations between ACC activity in conflict trials and PFC activity in the following trials (e.g. Kerns et al., 2004). At the same time, the theory has been elaborated and refined by the incorporation of learning processes to explain specific congruency proportion effects (Blais, Robidoux, Evan, & Besner, 2007; Verguts & Notebaert, 2008), conflicts at different levels of processing (Davelaar, 2008), and distinctions between reactive and proactive control to explain even complex data patterns from task switching studies (Brown, Reynolds, & Braver, 2007; Goschke, 2000). Notably, these model extensions were based on the fine-tuning of the basic conflict monitoring model (e.g. Davelaar, 2008; Verguts & Notebaert, 2008).

While these efforts advanced the theoretical integration of empirical findings, two alternative lines of thought received much less attention. One line questions the necessity of a specialized conflict monitoring module for the resolution of conflict and adjustments of control (Mayr & Awh, 2009; Ward & Ward, 2006). Instead of specific modules dedicated to certain cognitive operations, this line emphasizes cognitive functions to emerge from network dynamics (cf. Bressler & Kelso, 2001). In support of such a notion, several functional imaging studies have questioned the necessity of ACC involvement in conflict tasks, hence questioning a tight link between conflict monitoring processes and ACC functioning (see Mansouri, Tanaka, & Buckley, 2009 for a review). The second line of thought questions the assumption that conflict adaptation reflects adjustments of control only *after* a conflict trial (in the inter trial interval or at the beginning of the next trial, based on the accumulated amount of conflict at the end of a conflict trial). Instead, this line assumes that in parallel with the accumulation of conflict within a conflict trial, conflict resolution processes lead to readjustments of control already *within* a conflict trial (e.g. Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002; Goschke & Dreisbach, 2008; Ridderinkhof, 2002; Scherbaum, Fischer, Dshemuchadse, & Goschke, 2011). These adjustments are then carried over to the next trial, leading to the effect of conflict adaptation effects. One prediction derived from this view is that, in the course of a conflict trial, attention should be focused more strongly on relevant information than the course of a non-conflict trial. This shift in attention would then be carried over to the next trial. Indeed, a recent continuous EEG study found this pattern in a Flanker task by analyzing frequency tagged EEG (Scherbaum et al., 2011). Another expectation is that conflict adaptation effects should become weaker when the inter trial interval is increased, since the carry-over of control adjustments decays. Indeed, this has been shown recently for large inter trial intervals (Egner, Ely, & Grinband, 2010).

While the theoretical importance of the first line seems obvious in light of efforts to pin down the neural correlates of the conflict monitoring module (e.g. Kerns et al., 2004), the second line may seem to concern a subtle point at first sight. However, the fact that extensions of conflict monitoring theory build exactly on the assumption of measuring conflict and initiating adaptation only at the end of a conflict trial (e.g. Verguts & Notebaert, 2008) changes this argument from a subtle difference in definition into a possibly important difference in model architecture.

3. Continuous conflict adaptation without conflict monitoring

In this article, we propose an alternative neural network model that integrates the two views with the original idea of conflict monitoring theory, namely that conflict is responsible for increases of cognitive control. However, instead of an explicit conflict detection and signaling module causing these increases, we propose that context-sensitive adjustments of cognitive control occur within trials (see also Mayr & Awh, 2009), as an emergent by-product of the interaction dynamics within the network. Hence, this model serves as a proof of concept to demonstrate how conflict resolution and adjustments of cognitive control within a trial could lead to behavioral effects of control adjustments at different time scales, namely the conflict adaptation effect from trial to trial and the congruency proportion effect across series of trials as originally explained by conflict monitoring theory. The model is based on three assumptions leading to specific requirements for the network's architecture.

First, in line with previous models (e.g. Cohen, Servan-Schreiber, & McClelland, 1992) we assume that the resolution of conflict within a trial results from the interaction of potentially relevant input information and currently active goal representations, biasing processing towards goal-related information. Hence, bottom-up and top-down processes are coupled directly, providing a balance between stable behavior and flexible behavior (cf. Goschke, 2003). This is accomplished by top-down biases from activated goals on the one hand and by bottom-up input influencing the activation of goals on the other hand (Gilbert & Shallice, 2002).

Second, we assume that the conflict adaptation effect also results from the interaction dynamics resolving conflict in the previous trial. From this view, the conflict adaptation effect is a carry-over of control adjustments in the previous trial, in particular the strengthened goal activation resulting from the time-consuming conflict resolution processes *within* the previous trial (e.g. Burle et al., 2002; Fischer, Dreisbach, & Goschke, 2008; Goschke & Dreisbach, 2008; Ridderinkhof, 2002; Scherbaum et al., 2011). Note that this is a central difference to conflict monitoring theory, which builds on the conflict monitor as a special memory of accumulated conflict from the previous trial to adjust control after the trial. By contrast, in our model there is no explicit memory trace coding the strength of previous conflict, but the network simply ends up in a different state following conflict resolution. This change in state accounts for conflict adaptation effects in subsequent trials. From this assumption, two properties of the simulated network can be derived. One property concerns goal activation dynamics in the network, reflecting the activation of cognitive control: Goals need to exhibit semi-stable attractor dynamics (see also Rolls, 2010; Thelen, Schöner, Scheier, & Smith, 2001). On the one hand, goals need to show relatively stable states of activation, on the other hand, the states should be able to vary to a certain degree. Taken together, this kind of a broad attractor basin would allow for goal stability on the one hand and variability in goal activation on the other hand, leading to flexible cognitive control. The other property concerns the operationalization of time in the simulation: the transitions between trials needs to be modeled continuously to include decay effects of goal activation over time (see Gilbert & Shallice, 2002 for approaching this issue with fixed decay terms between trials; see Egner et al., 2010 for decay effects of control adjustments from trial to trial).

Third, we assume that conflict adaptation effects and effects of the relative frequencies of congruent and incongruent trials in a task (congruency proportion effect) result from consecutive network states reflecting an amplified or attenuated control over consecutive previous trials. To allow this integration over time, we follow the proposition of a time scale gradient on the posterior/anterior axis of the brain, with posterior sensory regions showing the shortest and anterior regions showing the longest time scale (Fuster, 2001; Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Kiebel, Daunizeau, & Friston, 2008). Analogous to this assumption, in our model units in the goal layer operate at a slower time scale than units in the input and response layers.

To validate the model, we will present a simulation comparing simulated RT data to classical human data, including conflict adaption and the congruency proportion effect as it has been shown by Gratton and colleagues (1992). Furthermore, we will evaluate the activation dynamics of the model within trials by a comparison to the activation dynamics found in a recent EEG study investigating within trial adjustments of control (Scherbaum et al., 2011). Finally, regarding the multitude of functions ascribed to the ACC (for reviews, see e.g. Botvinick, 2007; Mansouri et al., 2009; Rushworth, Walton, Kennerley, & Bannerman, 2004), we will evaluate how correlations of BOLD signals between ACC and PFC could be reinterpreted to offer a possible explanation for the apparent relationship between the two structures.

4. Simulation

4.1. Model and hypotheses

4.1.1. Layers

An outline of our model can be seen in figure 1 A (for details, please see Appendix I). Similarly to the structure of previous models solving conflict tasks (Botvinick et al., 2001; Cohen et al., 1992; Gilbert & Shallice, 2002), the model contains two input layers, a response layer and a task/goal layer.

-- Figure 1 --

The units in the input layers represent the relevant and irrelevant information in a conflict task, e.g. color and word meaning in a Stroop task. The output layer contains units indicating the response of the model, e.g. left or right in a classical conflict paradigm. Finally, the task/goal layer contains units representing the currently active task goal, e.g. color naming or word reading in a Stroop task.

4.1.2. Connectivity

Similarly to the connectivity of previous models, the input layers are connected via feed-forward connections to the output layer. To capture the fact that typical conflict paradigms are usually designed in a way that one type of information triggers habitual responses (e.g. word reading in the Stroop task), can be processed faster (e.g. location information in the Simon task), or is more dominant (e.g. 4 flanker stimuli vs. one target in the Flanker task), the input layer representing the stronger type of information has a stronger connection to the response layer than the input layer representing the weaker type. This difference in connection weights serves to simulate the a-priori advantage for the stronger information.

Different from typical conflict monitoring models, the goal layer and the input layers are connected via bidirectional recurrent connections (compare e.g. Gilbert & Shallice, 2002). Considering the connections from the input layers to the goal layer (the bottom-up direction), each input layer (e.g. the layer representing word color) excites its respective goal (e.g. color naming) and inhibits the opposite goal. Considering the connections from the goal layer to the input layer (the top-down direction), the unit representing one task goal (e.g. color naming) excites the respective input layer (word color) and inhibits the other input layer (e.g. word meaning). The top-down connections from the goal to the input layers enable an attentional bias of information processing that is typical for previous models (Botvinick et al., 2001; Cohen et al., 1992; Gilbert & Shallice, 2002). If one task is active and hence the respective goal unit is active, this unit will strengthen the activation of all input units representing task-relevant information (e.g. color) in the input layer, thereby biasing the competition between units representing task-relevant information (e.g. color) and units, representing task-irrelevant information (e.g. word meaning) in favor of the task-relevant information. However, the bottom-up connections allow stimulus information to influence the strength of the task goals. Hence, this connection can be used to activate a goal, to modify the activation strength of a goal by varying its supporting input, or even to 'capture attention' by activating a new goal using a very salient stimulus (compare Gilbert & Shallice, 2002).

4.1.3. Activation dynamics

While the pattern of connectivity in the network defines the static architecture of the model, the unit's activation dynamics define its reaction to input. Following previous work on cognitive control (Cohen et al., 1992) and working memory dynamics (e.g. Spencer & Schöner, 2003), a non-linear sigmoid activation function is applied on each unit's activation level (see figure 1B and Appendix I). This ensures that each unit participates in the interaction between units only to the extent that its activation exceeds a soft threshold modeled by the sigmoid function (e.g. Erlhagen & Schöner, 2002). In combination with recurrent excitatory connections of each unit, the resting level activation of each unit defines its activation state stability.

We used the non-linear properties of the sigmoid activation function to implement the required semi-stable goal attractor dynamics. Hence, goal units exhibit multi-stable dynamics (see figure 1B) with one stable state (activation at resting level = goal OFF/ deactivated) and one semi-stable state (reverberating activation[1] = goal ON/ activated). Without further input, goal units stay in the attractor basin of their current state. However, external influences can modulate and, if strong enough, switch the goal activation state. On the one hand, activating input, e.g. from goal congruent units in the input layer, drives a goal unit into the ON state (note, that this semi-stable state can then be maintained for a long time even in the absence of any input). On the other hand, inhibiting input, e.g. from goal incongruent units in the input layer, drives a goal unit into the OFF state if a certain level of deactivation is reached (note, that this state will be stable in the absence of any further excitatory input).

Since both goal units are coupled by inhibitory connections, together they form a goal attractor network with three stable attractor states: both goals deactivated, only goal 1 activated, only goal 2 activated. Hence, activation of one goal unit, e.g. by respective input or a task-cue at the start of a trial, can cause this unit to stay robustly activated in the absence of further input or even in the presence of input to the competing goal unit (see figure 1C). Furthermore, if one goal is in the ON state, it inhibits its opponent, suppressing any input to this other unit and keeping it in the OFF state. Due to the sigmoid activation function, as long as one goal is in the ON state, input sent to the opponent goal unit is relatively ineffective (one could even say ‘lost’), unless it is strong enough to overcome the inhibition and the soft, sigmoid threshold, driving the deactivated unit into the ON state while driving the originally activated unit into the OFF state. This could then be compared to a ‘capture of attention’ by a very salient input (compare Gilbert & Shallice, 2002). Taken together, the goal attractor network exhibits semi-stable activation dynamics. On the one hand, activation states are caught within the respective attractor basins (goal ON or goal OFF). On the other hand, a modulation of activation within the attractor basin is still possible (comparable to fluctuations in the activation stability of neural assemblies).

In the following simulation we hypothesize that the modulation of goal activation within the goal attractor network will reflect conflict at different time scales (within trials, from trial to trial, and across several trials). Within trials, we expect conflict to increase the activation of the correct goal via the interaction of bottom-up and top-down connections. An important factor in this increase should be *time* by itself, since conflict trials need more time to resolve the conflict and hence leave more time for the interaction dynamics to play their role in increasing the correct goal and enhancing the contrast between relevant and irrelevant information. From trial to trial, the increased activation of the correct goal in conflict trials will be expressed behaviorally as conflict adaptation. Finally, across trials the accumulated increase of activation of the correct goal after several conflict trials will be expressed in the congruency proportion effect. The model parameters will be chosen to qualitatively match the data from Gratton and colleagues (1992). The parameters will be described in detail in Appendix I.

4.2. Simulated paradigm

A conflict paradigm was implemented in which simulated participants responded to congruent and incongruent input information. A congruent trial meant that both input layers received the same input, e.g. [1,0] for stimuli indicating a left response. An incongruent trial meant that the two input layers received contradicting inputs, e.g. [0,1] and [1,0]. Trial to trial transitions were pseudo-randomized, balanced with respect to congruency (congruent/incongruent) and correct response (left/right). Model time was measured in cycles representing the time-steps at which activation was updated.

At the start of each simulated experiment, the goal to respond to the information in the relevant input layer was activated by feeding a neutral input pattern ([1,1]) to this input layer, simulating the instruction to respond to the respective information during the experiment. Note that this is possible because of the bottom-up connections between the input and the goal layer, activating the respective goal for the relevant input layer. The ‘instruction’ was shown only at the start of the experiment, before the start of the first trial[2]. The procedure for all following trials began with an inter-trial-interval showing no input (50 cycles). At the end of the inter trial interval, before the start of the next trial, all input units were shortly pre-activated (12 cycles) applying a neutral input pattern, simulating general response preparation (cf. Botvinick et al., 2001; Verguts & Notebaert, 2008).

To investigate the integration of congruency information at intermediate time scales (reflected in the congruency proportion effect), we performed runs with 80% congruent/ 20% incongruent trials, runs with 50% congruent/ 50% incongruent trials, and runs with 20% congruent/ 80% incongruent trials. For each congruency proportion condition, 10 simulated participants performed 256 (80/20), 240 (50/50), or 256 (20/80) trials per run.

4.3. Hypotheses

As a proof of concept, we expected the model to be able to show typical congruency, conflict adaptation, and congruency proportion effects in the simulated RTs as found by Gratton and colleagues and as simulated by the original conflict monitoring model.

With respect to model dynamics, we had two hypotheses. First, we hypothesized that goal activation will vary in dependence of the type of the presented trial (conflict, no conflict), especially at the end of the trial since longer reaction times leave more time for the interaction dynamics to push the activation of the correct goal in service of solving occurring conflict. Second, while the model configuration was chosen to match human RT data, we expected the model to also mirror human EEG data from a recent study (Scherbaum et al., 2011) investigating the continuous adjustments of attention to relevant and irrelevant stimuli in the face of conflict using frequency tagged EEG (Fuchs, Andersen, Gruber, & Müller, 2008; Müller, Andersen, & Keil, 2007; Müller, Teder-Sälejärvi, & Hillyard, 1998). Finally, we expected to find a correlation between the activation of goal units and the co-activation of response units to develop a possible explanation for the apparent relationship between conflict related activation and modulations of control.

4.4. Results

4.4.1. Simulated RT

As originally found in behavioral studies (Gratton et al., 1992; Ullsperger, Bylsma, & Botvinick, 2005), the model exhibited standard congruency and conflict adaptation effects (figure 2A), similar to the experiment 1 by Gratton and colleagues (1992): incongruent trials showed greater RT than congruent trials, but this difference was reduced for consecutive incongruent trials. Furthermore, the model exhibited a congruency proportion effect, showing stronger congruency effects in case of a higher frequency of congruent trials (figure 2B), similar to experiment 2 by Gratton and colleagues. Together, these RT data seem to confirm the ability of the model to integrate congruency information over time scales longer than single trials (for error data, see appendix II).

-- Figure 2 --

4.4.2. Goal units activation dynamics

To understand how the model performs conflict induced adjustments of control without using an explicit conflict monitoring module, we analyzed the activation dynamics of the goal unit representing the correct goal. Remember that the two goal units, inhibiting each other, constitute a stable goal attractor network with the three stable states: goal 1 activated, goal 2 activated, or no goal activated.

Dependent on the type of the previous trial, the activation of the goal unit already shows a difference, with trials following previously incongruent trials showing a higher goal activation than trials following previously congruent trials (figure 3A). At the time of response (marked in figure 3A by a cross), the activation additionally differed in dependence on the current trial type, with incongruent trials showing a higher activation than congruent trials. Notably, the difference between congruent and incongruent trials develops at the end of the trial, when incongruent trials have more time to further increase goal activation to resolve the occurring conflict. Furthermore, for all trials the activation of the goal unit representing the incorrect goal never surpassed the soft sigmoid threshold due to inhibition by the correct goal. Hence, the incorrect unit stayed deactivated despite supporting input coming from the input layer.

Taken together, this indicates two points. First, the model retains within the goal layer information about previously occurred conflict and current conflict, accounting for the conflict adaptation effect in RT data. Second, the longer time it takes to settle on a response decision in incongruent trials enables the interaction dynamics to strengthen the activation of the correct goal supporting conflict resolution.

Since the model also showed a congruency proportion effect, we analyzed the activation dynamics for the different conditions of congruency proportion. We focused on incongruent trials following congruent trials, since we expected these trials to be influenced most strongly by congruency proportion. Compared to the mean dynamics, simulations with a low proportion of incongruent trials showed a weaker activation of the correct goal at the beginning of the trial as compared to simulations with an equal or high proportion of incongruent trials (figure 3B).While this difference weakens within the trial, it is still present at the end of the trial. This indicates that the model integrates information about conflict across several trials within the goal layer

Taken together, the activation dynamics of the goal layer show that information about congruency on different time scales is reflected in the different activation states of the correct task goal. While conflict leads to strengthening of the correct task goal within a trial and hence strengthens the focusing of attention on relevant information in the service of conflict resolution, this stronger activation is carried over to following trials, leading to conflict adaptation from trial to trial and congruency proportion effects across trials. This indicates that conflict resolution and conflict adaptation interact within the goal layer and are not necessarily functionally separated from each other.

4.4.3. Input layer dynamics

On the input layer of the model, the continuous adjustments of control found for the goal layer lead to a continuous enhancement of the activation difference between relevant and irrelevant information. Hence, the difference between congruent and incongruent trials for the activated input units (see figure 4, left) reveals that the relevant information was amplified over the course of a conflict trial while the irrelevant information was attenuated. This contrast enhancement served to support conflict resolution within the conflict trial itself. Even more important, this contrast enhancement was stronger for trials following a previous congruent trial, when the system was not well prepared for conflict, compared to trials following a previous incongruent trial, when the system was already prepared for conflict (see figure 4, left, inset). This pattern is in concordance with previous studies investigating neural activation patterns across trials (e.g. Egner & Hirsch, 2005).

-- Figure 4 --

In a recent study, we used EEG frequency tagging to investigate exactly this kind of within-trial contrast enhancement (Scherbaum et al., 2011). In a flanker task, relevant and irrelevant stimuli were tagged with different flicker frequencies. The amplitudes of the resulting oscillatory signals in the EEG indicated the amount of attention allocated to the different stimuli (Fuchs et al., 2008; Müller et al., 2007, 1998) and showed dynamics similar to the ones exhibited by the model (compare figure 4, right), including the difference between trials preceded by different types of previous trials (figure 4, left, inset). Notably, this similarity was found despite the model fitting being restricted to behavioral RT data. We interpret this as support for the validity of our basic assumptions underlying the model even on the neural level.

4.4.4. Correlations of response and goal activation

The final open question concerned the relationship of response conflict and goal activation, as it has been illustrated by correlations between the BOLD signal in the ACC and the BOLD signals in the PFC on subsequent trials (Kerns et al., 2004). How could such a correlation be explained within the framework of the presented model? Considering the diversity of cognitive functions related to ACC activity (e.g. Mansouri et al., 2009; Rushworth et al., 2004), a first, yet speculative attempt to integrate these findings into the presented framework rests on the assumption of inhibitory feed-forward activation as one determinant of ACC engagement (for the debate about this issue, see e.g. Lavric, Pizzagalli, & Forstmeier, 2004; Nieuwenhuis, Yeung, Van Den Wildenberg, & Ridderinkhof, 2003). In the model, we operationalized this as the activation of inhibitory inter-neurons performing lateral inhibition between the different neural assemblies representing the possible responses[4]. This activation could closely resemble the original conflict monitoring signal since it reflects a linear combination of the parallel activation of different responses[3] (Botvinick et al., 2001), however without any causal connection to the activation of the goal units. To validate this interpretation in the context of our model, we implemented a similar analytic approach to our simulated data as has been applied previously to BOLD activation patterns in conflict tasks (Kerns et al., 2004). Specifically, we first computed correlations between the summed activation of the response units on the current trial. This activation should reflect the overall inhibitory activation within the response layer and hence the hypothesized activation of the ACC. Second, we calculated the activation of the goal unit on the following trial. Similarly to original conflict monitoring theory and following the interpretation of Kerns and colleagues, we assume that goal activation is represented by neural assemblies in the lateral prefrontal cortex (LPFC). Third and finally, we correlated these two values, hypothesized activation of the ACC in trial N, and the hypothesized activation of the LPFC in trial N+1.

-- Figure 5 --

As shown in figure 5, the resulting positive correlation (r^2 = 0.89) closely resembles the results found by the original imaging study (Kerns et al., 2004). Most importantly, however, the model produced this correlation even though the response and the goal layers are not directly interconnected. Notably, this correlation could be found for the activity integrated across time as it is also reflected in fMRI signals, indicating the similarity between the classical conflict signal and the inhibitory activation calculated here. However, this similarity to the conflict signal changes when considering activation peaks (instead of integrated activation) as they might be measured by EEG recordings, e.g. the N2 (e.g. Yeung, Botvinick, & Cohen, 2004; but see also Burle, Roger, Allain, Vidal, & Hasbroucq, 2008). While the peak of the conflict signal was sensitive to congruency proportion as expected from previous studies (Yeung et al., 2004), there was no general difference in peak size between congruent and incongruent trials: since the assumed inhibitory activation reflects the summed activation of both responses, the peak of the inhibitory activation did not differ between congruent and incongruent trials, distinguishing it from predictions of a classical conflict signal as it might be measured by fMRI (as also pointed to by Burle et al., 2008).

5. Discussion

We presented a dynamic connectionist model as an account of how cognitive control could be adjusted continuously within trials and without building on an explicit conflict monitoring module. As a proof of concept, the model successfully reproduces standard effects indicating the adaptation of control across trials, i.e. conflict adaptation and the proportion congruent effect. Critically, the model was able to account for these effects without the assumption of an explicit conflict monitoring module. Rather, context-sensitive adaptations of control across trials emerged as a by-product of the fact that increased RT on conflict trials gave rise to a different within-trial activation dynamics.

In the simulation, the model was able to adjust control to conflict within a trial, leading to effects at larger time scales, namely conflict adaptation and congruency proportion effects. Furthermore, the model mirrored EEG data showing control adjustments within conflict trials, despite the fact, that the model was not fitted to this continuous data. The model also reproduced effects of conflict adaptation at the input level as shown in previous imaging studies examining conflict adaptation across trials (Egner & Hirsch, 2005) and as shown in a recent EEG study focusing on within-trial control adjustments (Scherbaum et al., 2011). This indicates that congruency effects, conflict adaptation, and the congruency proportion effect as simulated by the original conflict monitoring model could stem from the same process: the continuous interaction between input information and goal representations in a multi-stable goal attractor network, resolving conflict and adapting control through time within a conflict trial. In contrast, conflict monitoring theory postulates that these effects stem from at least two processes, conflict resolution within a trial and adjustments of control across trials (as reflected in the conflict adaptation effect and the congruency proportion effect). These two processes are separated in space (*resolution* performed by the main network and *adjustments* performed by the conflict monitoring module) and time (*resolution* within the trial and *adjustments* across trials).

Addressing the separation in space, our model indicates, as a proof of concept, that the adjustment of cognitive control could be implemented without the need of a distinct conflict monitoring module evaluating the requirement for control (Botvinick et al., 2001). This line of thinking has shown up previously in the literature (e.g. Mayr & Awh, 2009; Ward & Ward, 2006), but has received little attention. While Ward & Ward (2006) were able to show that conflict resolution is possible without explicit conflict monitoring (similarly to the model of Cohen et al., 1992 that we extended here), there has been no model addressing mechanisms of conflict driven adjustments leading to conflict adaptation without conflict monitoring. However, two lines of arguments suggest reconsidering the conceptualization of the conflict monitor as it can be found in conflict monitoring theory. First, the discussion questioning a strict modularity of mind (Fodor, 2005, see e.g. 1983, 2001; Pinker, 1997, 2005a, 2005b) and the dynamic approach to cognition (e.g. Bressler & Kelso, 2001; Kelso, 1995; Scherbaum, Dshemuchadse, & Kalis, 2008; Thelen et al., 2001) both promote a view focusing on the emergence of cognitive function from the interaction dynamics within the system. Hence, instead of a specific mapping of fine grained function (e.g. conflict monitoring) to structure and specific modules in the neural system, we followed a more general approach assuming that there is a hierarchy of timescales in the brain (Fuster, 2001; Hasson et al., 2008; Kiebel et al., 2008) that leads to a certain degree of functional specialization, e.g. the processing of stimulus properties in the input layer and the maintenance of information in the goal layer. Second, the conflict detector in conflict monitoring explicitly calculates the activation energy within the response layer as conflict signal, an operation that is not a neural computation in the classical sense but rather a mathematical measure of activation state stability[3]. From the perspective of biological plausibility, performing this explicit calculation is at least questionable. Within the original framework, the importance of both objections might be a matter of definition and discussion. However, the necessity for biologically plausible mechanisms becomes more obvious, when extensions build explicitly on these critical assumptions of the original model. For example, the model proposed by Verguts and Notebaert (2008) uses the conflict signal, calculated at the end of each trial, as a quantity that modulates Hebbian learning in their model. Hence, in that model, the hypothesized conflict signal becomes a decisive part used for the working of a core feature of that model. It becomes by far more than a matter of discussion, since changing the nature of the conflict signal now changes all the parts building on its exact specification.

Addressing the separation in time, our model lends further credibility to the assumption that the process of conflict resolution includes adjustments of control within a conflict trial. These adjustments within the trial cause the conflict adaptation effect across trials by being carried over from the previous conflict trial to the next trial. While the timing issue of conflict adaptation has still not yet been resolved, most extensions of conflict monitoring theory followed the original approach of measuring response conflict only at the end of each trial - information that is used afterwards to adjust control according to the just experienced level of conflict. However, several authors argue for a different version of conflict induced adjustments (Fischer et al., 2008; Goschke, 2003; Mayr & Awh, 2009; Scherbaum et al., 2011) with the occurrence of conflict leading to adjustments within the conflict trial itself. Similar to the objections against the spatial separation of conflict resolution and conflict driven adjustments of control, the difference seems to be subtle at first. However, the difference becomes decisive, when extensions of conflict monitoring theory build on the specific assumption of control adjustments only after a conflict trial. Again, a prominent example is the model proposed by Verguts and Notebaert (2008) that uses the conflict signal only at the end of the trial to modulate Hebbian learning. Since this weight modification is done at the end of the trial, this step strongly builds on the properties of the conflict signal at the end of the trial. Hence, the exact timing of when conflict is signaled now plays a decisive role in that model and is not only a matter of discussion any more.

Our model unites these processes, conflict resolution and conflict driven adjustments, by using differences in trial duration, caused by response conflict, to allow the network interaction dynamics to adjust the network state and resolve conflict. If a distinct conflict-monitoring process is not necessary as suggested by our model, how could several imaging studies provide evidence for the supposed neural conflict monitoring module, located in the anterior cingulate cortex (ACC; see e.g. Botvinick et al., 2004; Kerns et al., 2004)? While these findings originally supported the neural underpinnings of conflict monitoring theory, other studies exist that shed at least some doubt on the decisive role of the ACC in conflict tasks (for a review, see e.g. Mansouri et al., 2009). Furthermore, the ACC has been related to many different functions in recent years, including error likelihood (Brown & Braver, 2005), number of responses (Brown, 2009), avoidance learning (Johansen & Fields, 2004), tracking uncertainty or environmental volatility (Rushworth & Behrens, 2008), or relating actions to their outcomes (Rushworth et al., 2004). While there are first approaches to integrate all these different views on ACC functioning (e.g. Botvinick, 2007), it seems difficult to formulate an integrated theory of ACC functioning (Mansouri et al., 2009). Further research is clearly necessary to clarify which cognitive function or functions may best characterize the role of the ACC. In this article, we proposed an alternative explanation for the observed correlation of ACC and PFC activity in imaging studies, by following a simple, yet speculative, assumption that feed-forward inhibitory processes rather than conflict-monitoring might be reflected in ACC activity (for the debate about this issue, see e.g. Lavric et al., 2004; Nieuwenhuis et al., 2003). Considering its suitable anatomical position, and considering that ACC activity is only found for response conflict but not for stimulus conflict (van Veen & Carter, 2002), one could interpret ACC engagement as the activation of pools of inhibitory inter-neurons performing inhibition between the different neural assemblies representing competing responses[4], if this is necessary. Following this interpretation, we calculated this ACC activation for our model by summing up the activation of all response units, since each response unit could send its activity to its competitor via the ACC. This summed activity in a conflict trial correlated with the activation of the relevant goal unit in the following trial, similar to the correlation of ACC and PFC activation in typical imaging studies (Kerns et al., 2004). While this correlation was previously interpreted in favor of the ACC's role in conflict monitoring and signaling, the model produced this correlation even though the response and the goal layers had no direct connection to each other. One may conclude that a positive correlation between inhibition-related ACC activity and subsequent PFC activity per se does not *necessarily* indicate a direct causal relationship between these brain areas, but could alternatively be the product of two independent processes. One of these processes, the inhibition of opponent responses reflected in ACC activation, might produce activation that seems like a conflict signal, but is in fact only an 'epiphenomenon' of inhibitory activity that could even be dispensable in specific conflict paradigms (Mansouri et al., 2009). This could explain why patients with lesions in the ACC (hosting the dispensable process) do not necessarily show behavioral impairments

(Fellows & Farah, 2005) while patients with lesions in the PFC (providing the necessary ‘control process’ due to its position in the hierarchy of timescales) do show impairments (e.g. Vendrell et al., 1995). Furthermore, it also offers an explanation for why ACC activation is not necessarily found in fMRI based conflict studies (Erickson et al., 2004; Milham, Banich, Claus, & Cohen, 2003). Finally, in our model, the difference in inhibitory activity integrated across time between congruent and incongruent trials mainly came from the difference in trial length and not from the difference in peak strength. However, signal peaks were still sensitive to congruency proportion as it has been show previously for EEG measures of conflict (Yeung et al., 2004). Hence, these results contribute to the diverse discussion about different neural correlates of conflict detection (Burle et al., 2008; Nieuwenhuis et al., 2003; Yeung et al., 2004): While fMRI integrates across time, leading to typical conflict related activation differences in dependence of trial length (compare e.g. Grinband et al., 2011; but see Yeung, Cohen, & Botvinick, 2011), EEG components, measuring peak activity, might not show these differences, as it has been indicated previously (e.g. Burle et al., 2008).

It must be emphasized that conflict monitoring theory took a big step by defining conflict as a driving force for the adaptation of cognitive control. In their original work, Botvinick and colleagues rendered their definition of conflict (energy within the response layer) as a simple, yet preliminary way of using conflict as a signal. Our model provides an alternative definition of how conflict drives adaptation, by prolonging trials and adding time for the interaction dynamics of the network to change goal activation states within the goal layer. Hence, our model preserves the function of conflict as a driving force for adaptation, while superseding the explicit signaling of conflict to adjust cognitive control[5]. Furthermore, it contributes a possible solution to the debate whether ACC activation might only be related to time on task (Grinband et al., 2011). In this debate, proponents of conflict monitoring theory emphasized the role of response conflict as a psychological explanation for the found phenomena of conflict adaptation (Yeung et al., 2011). With trial timing at its core, the presented model builds on the necessity of response conflict as a factor prolonging trials and offers, at the same time, a mechanism how the conflict induced prolongation of trials could lead to the adaptation of control across trials without the necessity of an explicit conflict monitor.

If the model uses additional time for the interaction dynamics of the network to change goal activation states within the goal layer, does this mean that the all effects of conflict could simply be reduced to be effects of trial length? Three reasons complicate such a straight forward interpretation. First, the model represents only a core part of the cognitive system. Hence, multiple possible causes leading to the differences in trial timing need to be considered. For example, in a mouse tracking study (Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), we found a large amount of slow trials showing wrong initial guessing. It is open, if such slow trials lead to more intense processing within the core part of the cognitive system. Second, even in congruent trials, conflict can occur occasionally (as it has been shown e.g. by EMG studies, see Burle et al., 2002) and hence, conflict monitoring theory could make a similar prediction, since these trials are prolonged by occurring conflict. Third, considering the diverse reasons for variance in trial timing, an empirical validation of this specific claim becomes complex, since one needs to manipulate trial length directly without increasing conflict. This might be done by increasing perceptual difficulty (see e.g. Dreisbach & Fischer, 2011),but care must be taken not to induce any kind of response conflict. While these points make the argument more complex, the need for such clarifications matches the aim of the current article: to propose an alternative view on conflict driven adjustments of control that inspires further research into this direction.

Similar to the original conflict monitoring theory, our model is presented as proof of concept and enhances the original approach in a different direction than recently presented modifications of the conflict monitoring model. Certainly, further work within the presented framework owes to provide the integration of further effects from the empirical literature. An important point concerns the influence of the proposed conflict adaptation mechanism on error rates. While our main simulation focused on modeling RT effects, data in appendix II indicates that the biasing top-down influence of the goal units also influences error rates in a similar and expectable way. Error rates are higher when irrelevant information has its strongest influence, in incongruent trials following congruent ones. In consequence, in incongruent trials following a previous incongruent trial, error rates are lower, since irrelevant information is attenuated and relevant information is amplified by increased correct goal activation. The same holds for the influence of the proportion of congruent trials (see appendix II). Another prominent example of further effects is the item-specific congruency proportion effect (for the respective models, see e.g. Blais et al., 2007; Verguts & Notebaert, 2008) which is defined as a different amount of adaptation over longer time scales for items with a different amount of congruent trials. While the presented work has not integrated this effect (similarly to original conflict monitoring theory, see Blais et al., 2007), the importance of trial timing hints to a possible, though yet speculative, future solution: Instead of applying Hebbian learning at the end of a trial, modulated by the amount of occurred conflict, continuous Hebbian learning could accumulate within a trial and should lead to similar results as it has been shown by Verguts and Notebeart (2008). A final prominent example is the integration of conflict at different timescales. The presented model was able to integrate this information within one goal unit and hence followed the original conflict monitoring model for explaining conflict adaptation and congruency proportion effects.

However, recent data (Funes, Lupiáñez, & Humphreys, 2010) indicate that control might be adapted across different timescales by different mechanisms (De Pisapia & Braver, 2006). While the presented model does not integrate these findings, yet, the assumption of a hierarchy of timescales in the brain (Fuster, 2001; Hasson et al., 2008; Kiebel et al., 2008) points to a plausible future extension within the presented framework, with different goal units working at different timescales.

A final important point is the stability of model behavior to parameter variations. The present model shows qualitatively stable behavior for different combinations of parameter settings, providing a corridor of valid combinations in the multidimensional parameter space. However, the model is not robust against several parameter variations, which lies in its dynamic nature for two reasons. First, in a dynamic model, parameters are not merely settings that are tuned to fit the data, but instead provide insight into the functioning of the model and links to variations in real behavior (see e.g. Rolls, Loh, & Deco, 2008; Schöner & Thelen, 2006; Thelen et al., 2001). For instance, variations of resting level and lateral inhibition in the goal layer can provide an approach to inter-individual differences in cognitive control (compare e.g. Johnson, Spencer, & Schoner, 2008; Thelen et al., 2001), mirroring e.g. developmental changes in structure or differences in the level of neurotransmitters. Hence, the models architecture becomes twofold, with the features of the "spatial" architecture of the model, including connections, number of layers, number of nodes, are not less parameters than the "functional" architecture, i.e. resting level, the strength of lateral inhibition, and the time scale. From this dynamic perspective, the model is defined on the one hand by its spatial architecture and on the other hand by its functional architecture (note that this is mirrored in figure 1, showing on the one hand the spatial architecture in figure 1A and on the other hand the functional architecture in figure 1B and 1C).

Second, our simulation was necessarily continuous over time and, hence, the impact of small parameter variations accumulates over time leading to varying results. This stands in contrast to other simulations that run trial-wise and reset values at the start of every trial, adding stability to the simulation. Notably, running a model continuously over time adds biological plausibility to the simulation, but at the cost of stability to variations in parameters. Taking these points together, in the current setup, the model is provided as a proof of concept and model behavior is susceptible to changes in parameters. However, it should be noted that while care has to be taken in choosing the parameters for a model, not every model will produce specific results if the parameters are tuned fine enough (see e.g. McClelland, 2009).

In conclusion, this article presented a new approach to the question how cognitive control is recruited flexibly according to task demands, based on network dynamics. We proposed that this is possible without the use of an explicit module performing conflict monitoring (Botvinick et al., 2001), by building on adjustments of control within conflict trials supporting conflict resolution. While we aimed to provide a first proof of concept with this model, further empirical and modeling work is certainly necessary to provide additional insight into the dynamic properties of the engaged processes and mechanisms of cognitive control and to fit the model to an increasing number of empirical results on the context-sensitive adjustments of cognitive control.

6. References

Blais, C., Robidoux, S., Evan, F. R., & Besner, D. B. (2007). Item-Specific Adaptation and the Conflict-Monitoring Hypothesis: A Computational Model. *Psychological Review*, *114*(4), 1076–1086.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 356.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Science*, *8*(12), 539–546.

Bressler, S. L., & Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, *5*(1), 26–36.

Brown, J. W. (2009). Conflict effects without conflict in anterior cingulate cortex: multiple response effects and context specific representations. *NeuroImage*, *47*(1), 334–341.

Brown, J. W., & Braver, T. S. (2005). Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science*, *307*(5712), 1118–1121.

Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology*, *55*(1), 37–85.

Burle, B., Possamaï, C. A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336.

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error Negativity Does Not Reflect Conflict: A Reappraisal of Conflict Monitoring and Anterior Cingulate Cortex Activity. *Journal of Cognitive Neuroscience*, *20*(9), 1637–1655.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American journal of psychology*, *105*(2), 239–269.

Davelaar, E. J. (2008). A computational study of conflict-monitoring at two levels of processing: Reaction time distributional analyses and hemodynamic responses. *Brain Research*, *1202*, 109–119.

Dreisbach, G., & Fischer, R. (2011). If it's hard to read… try harder! Processing fluency as signal for effort adjustments. *Psychological Research*, *75*(5), 376–383.

Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: characterizing the time-course of congruency sequence effects. *Frontiers in Cognition*, *1*, 154.

Egner, T., & Hirsch, J. (2005). The neural correlates and functional integration of cognitive control in a Stroop task. *Neuroimage*, *24*(2), 539–47.

Erickson, K. I., Milham, M. P., Colcombe, S. J., Kramer, A. F., Banich, M. T., Webb, A., et al. (2004). Behavioral conflict, anterior cingulate cortex, and experiment duration: Implications of diverging data. *Human Brain Mapping*, *21*(2), 98–107.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, *109*(3), 545–572.

Fellows, L. K., & Farah, M. J. (2005). Is anterior cingulate cortex necessary for cognitive control? *Brain*, *128*(4), 788–796.

Fischer, R., Dreisbach, G., & Goschke, T. (2008). Context-sensitive adjustments of cognitive control: conflict-adaptation effects are modulated by processing demands of the ongoing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 712–718.

Fodor, J. (2005). Reply to Steven Pinker "So How Does The Mind Work?" *Mind & Language*, *20*(1), 25–32.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (2001). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Fuchs, S., Andersen, S. K., Gruber, T., & Müller, M. M. (2008). Attentional bias of competitive interactions in neuronal networks of early visual processing in the human brain. *Neuroimage*, *41*(3), 1086–1101.

Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Sustained vs. transient cognitive control: Evidence of a behavioral dissociation. *Cognition*, *114*(3), 338–347.

Fuster, J. M. (2001). The prefrontal cortex - an update: time is of the essence. *Neuron*, *30*(2), 319–333.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive psychology*, *44*(3), 297–337.

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 331–355). Cambridge, MA: MIT Press.

Goschke, T. (2003). Voluntary action and cognitive control from a cognitive neuroscience perspective. *Voluntary action: Brains, minds, and sociality.* (pp. 49–85). Oxford: Oxford University Press.

Goschke, T., & Dreisbach, G. (2008). Conflict-triggered goal shielding: Response conflicts attenuate background-monitoring for concurrent prospective memory cues. *Psychological Science*, *19*(1), 25–32.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *NeuroImage*, *57*(2), 303–311.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *The Journal of Neuroscience*, *28*(10), 2539 – 2550.

Johansen, J. P., & Fields, H. L. (2004). Glutamatergic activation of anterior cingulate cortex produces an aversive teaching signal. *Nature Neuroscience*, *7*(4), 398–403.

Johnson, J. S., Spencer, J. P., & Schoner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology*, *26*(2), 227–251.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science, 303*(5660), 1023–1026.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11).

Lavric, A., Pizzagalli, D. A., & Forstmeier, S. (2004). When’go’and’nogo’are equally frequent: ERP components and cortical tomography. *European Journal of Neuroscience, 20*(9), 2483–2488.

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience*, *10*(2), 141–152.

Mayr, U., & Awh, E. (2009). The elusive link between conflict and conflict adaptation. *Psychological Research*, *73*(6), 794–802.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38.

Milham, M. P., Banich, M. T., Claus, E. D., & Cohen, N. J. (2003). Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage, 18*(2), 483–493.

Müller, M. M., Andersen, S. K., & Keil, A. (2007). Time Course of Competition for Visual Processing Resources between Emotional Pictures and Foreground Task. *Cerebral Cortex*, *18*(8), 1892–1899.

Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*(7), 631–634.

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, and Behavioral Neuroscience*, *3*(1), 17–26.

Pinker, S. (1997). *How the mind works*. New York: W.W. Norton & Co.

Pinker, S. (2005a). A Reply to Jerry Fodor on How the Mind Works. *Mind & Language*, *20*(1), 33–38.

Pinker, S. (2005b). So How Does the Mind Work? *Mind & Language*, *20*(1), 1–24.

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*(10–12), 1322–1326.

Ridderinkhof, K. R. (2002). Micro-and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, *66*(4), 312–323.

Rolls, E. T. (2010). Attractor networks. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 119–134.

Rolls, E. T., Loh, M., & Deco, G. (2008). An attractor hypothesis of obsessive-compulsive disorder. *European Journal of Neuroscience*, *28*(4), 782–793.

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, *8*(9), 410–417.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*(3), 407–416.

Scherbaum, S., Dshemuchadse, M., & Kalis, A. (2008). Making decisions with a continuous mind. *Cognitive, Affective, & Behavioral Neuroscience, 8*(4), 454–474.

Scherbaum, S., Fischer, R., Dshemuchadse, M., & Goschke, T. (2011). The dynamics of cognitive control: Evidence for within ‐ trial conflict adaptation from frequency ‐ tagged EEG. *Psychophysiology, 48*(5), 591–600.

Schöner, G., & Thelen, E. (2006). Using Dynamic Field Theory to Rethink Infant Habituation. *Psychological Review*, *113*(2), 273–299.

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, *6*(4), 392–412.

Stroop, J. R. (1935). Studies of interference in serial verbal interactions. *Journal of Experimental Psychology*, *18*, 643–662.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*(1), 1–34.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*(6), 727–735.

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict-adaptation effect: it's not just priming. *Cognitive, Affective, and Behavioral Neuroscience*, *5*, 467–472.

van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*(4–5), 477–482.

Vendrell, P., Junqué, C., Pujol, J., Jurado, M. A., Molet, J., & Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia, 33*(3), 341–352.

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological Review*, *115*(2), 518–525.

Ward, R., & Ward, R. (2006). Cognitive conflict without explicit conflict monitoring in a dynamical agent. *Neural Networks*, *19*(9), 1430–1436.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review, 111*(4), 931–959.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2011). Errors of interpretation and modeling: a reply to Grinband et al. *NeuroImage*, *57*(2), 316–319.

7.

Author Note

This research was supported by the German Research Foundation (DFG) (Grant Go720/3-3).

Footnotes

[1] This behaviour closely resembles the dynamics of neural activation in the LPFC in the absence of external input (for a review, see e.g. Curtis & D'Esposito, 2003).

[2] Instead of using a neutral input to activate the correct goal, it would also be possible to preset the correct goal as it is done in most conflict simulations.

[3] Note that the original conflict signal multiplied concurring response activation reflecting the energy potential within the response layer (Botvinick, Braver, Barch, Carter, & Cohen, 2001). While this is a practical mathematical shortcut, it does not follow the usual way of simulating neural activity by inputs transformed by a biologically plausible activation function.

[4] Note that in neural networks, units can send inhibitory and excitatory signals directly to other units, dependent on the connection weights. In contrast, in neural systems, excitatory neurons need to perform inhibition via the activation of inhibitory inter-neurons.

[5] This also distinguishes the presented approach from a recent proposal that conflict is detected and maintained in the prefrontal cortex (Mansouri, Tanaka, & Buckley, 2009). While the presented work supports the decisive role of the PFC (presumably representing the goals as they were represented in the model) in control adjustments and conflict resolution, it objects to the explicit detection of conflict and proposes an implicit effect on conflict on goal activation.

Appendix I

The model consists of 2 input layers, a response layer, and a goal layer, with 2 units per layer. Activation of each unit is calculated by non-linear first order differential equations as been used previously for patterns of neural activation (Amari, 1977; Erlhagen & Schöner, 2002). Simulated by numerical integration, results were obtained using Matlab 2006a running under Windows XP SP3.

The difference equation over time *t* for the activation *u* of units in a layer followed in principle this scheme:

$$\tau\dot{u}(t) = -u(t) - h + wi \cdot \sigma(u(t)) + w \cdot Input(t)$$

Here, *h* denominates the restlevel, *wi* the interaction weight within the layer (self excitation and lateral inhibition), *w* the weight of inputs into the layer, *Input* defines the Input into layer, and σ denotes the sigmoid non-linearity:

$$\sigma(x) = 1/(1 + e(-\beta \cdot (x - \alpha)))$$

Hence, each unit contributes to interaction in the network only to the extent that its activation exceeds a soft threshold (Cohen et al., 1992; Erlhagen & Schöner, 2002).

Following this scheme, the equation for the input layers was:

$$\tau\dot{u}(t) = -u(t) - h + wi_i \cdot \sigma(u(t)) + w_{si} \cdot S(t) + w_{gi} \cdot G(t)$$

Here, wi_i denotes the interaction weight within the input layers, w_{si} the weight of environmental inputs into the layers, w_{gi} the strength of bias from goal layer, *S* represents the environmental stimulation of the layer (0 for no stimulation, 1 for stimulation), *G* represents the bias from the goal layer, and σ denotes the sigmoid non-linearity defined above.

The very similar equation for the response layer is:

$$\tau\dot{u}(t) = -u(t) - h + wi_r \cdot \sigma(u(t)) + w_{i_1r} \cdot I_1(t) + w_{i_2r} \cdot I_2(t)$$

Here, wi_r denotes the interaction weight within the response layer, w_{i1r} and w_{i2r} denote the strength of input from the input layers, and *I1 and I2* represent the signal from the input layers. Responses were judged as given when σ ($u(t)$) reached a threshold of 0.9.

To implement the goal layer as multi-stable attractor network, this layer contained units with a bi-stable attractor layout and slower integration of input over time (see also Figure 1 B and C). This was implemented by choosing a different resting level (cf. Spencer & Schöner, 2003) and time scale:

$$\tau_g \dot{u}(t) = -u(t) - h_g + wi_g \cdot \sigma(u(t)) + w_{i_1 g} \cdot I_1(t) + w_{i_2 g} \cdot I_2(t)$$

Here, wi_g denotes the interaction weight within the input layer, w_{i1g} and w_{i2g} the strength of inputs from the input layers, *I1* and *I2* denote the inputs from the input layers.

Hence, we had a feed-forward network from input to response units, with processing in the input layer modulated by reciprocal connections to and from the goal layer.

The weight matrices are shown in the following. The interactions within the input, response, and the goal layer were defined by

$$wi_i = \begin{pmatrix} 0 & \text{-}0.5 \\ \text{-}0.5 & 0 \end{pmatrix}, \quad wi_r = \begin{pmatrix} 0 & \text{-}0.5 \\ \text{-}0.5 & 0 \end{pmatrix}, \text{ and } \quad wi_g = \begin{pmatrix} 0.8 & \text{-}5 \\ \text{-}5 & 0.8 \end{pmatrix}.$$

Hence, within the goal layer, there was strong lateral inhibition, while this parameter was equal for input and response layers.

Signal transmission from the input layers to the response layer was defined by

$$w_{i_1 r} = \begin{pmatrix} 14 & 0 \\ 0 & 14 \end{pmatrix}, \quad w_{i_2 r} = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}.$$

Hence, the input layer processing irrelevant information (input layer 1), was stronger associated with the response layer (representing the more habitual response bias, e.g. word reading in the Stroop task).

Signal transmission from the input layers to the goal layer and the goal layer to the input layers was defined by

$$w_{i_1 g} = \begin{pmatrix} 3.5 & -3.5 \\ 3.5 & -3.5 \end{pmatrix}, \quad w_{i_2 g} = \begin{pmatrix} -3.5 & 3.5 \\ -3.5 & 3.5 \end{pmatrix}, \quad w_{g i_1} = \begin{pmatrix} 1.5 & 1.5 \\ -1.5 & -1.5 \end{pmatrix}, \quad w_{g i_2} = \begin{pmatrix} -1.5 & -1.5 \\ 1.5 & 1.5 \end{pmatrix}.$$

Hence, each input layer strongly stimulates the supporting unit and inhibits the opposite unit in the goal layer, while the goal layer, in turn, biases competition in favor of its related input layer.

The other parameters where chosen as follows: h = 5, τ = 10, h_g = 0.95, τ_g = 40, w_{si} = 6, α = 0, β = 2. The parameters were chosen to match the RT data of Gratton and colleagues (1992), as they have been replicated by Botvinick and colleagues (2001). The qualitative pattern of results was robust for a certain range of parameter combinations.

Appendix II

To investigate the impact of errors, we rerun the simulation with the same parameters as in the original run (30 participants, 3 congruency proportions). Additionally, we added noise to the input and the response units (normally distributed noise with *M* = 0 and *SD* = 0.2).

As expected, the pattern of error rates mirrored RT data (see figure AII.1), with an increased error rate for incongruent trials following congruent ones compared to incongruent trials following incongruent ones. Furthermore, and also as expected, the size of this difference depended on the proportion of congruent trials, with less errors in incongruent trials in the high conflict condition (20% congruent trials) and more errors in the low conflict condition (80% congruent trials).

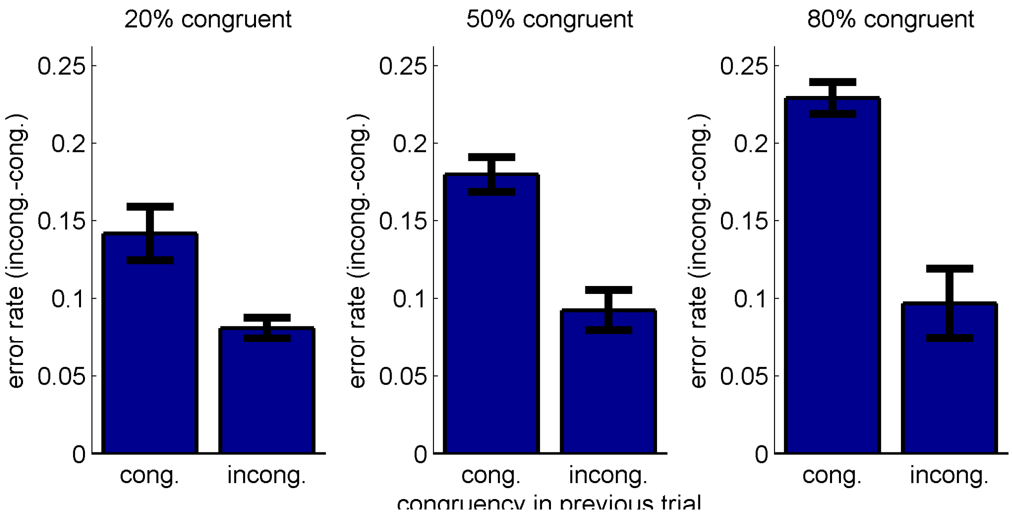


Figure AII.1. Impact of congruency on error rate (incongruent – congruent trials). Error rate is influenced by both, conflict adaptation and congruency proportion.

Figures

Figure 1

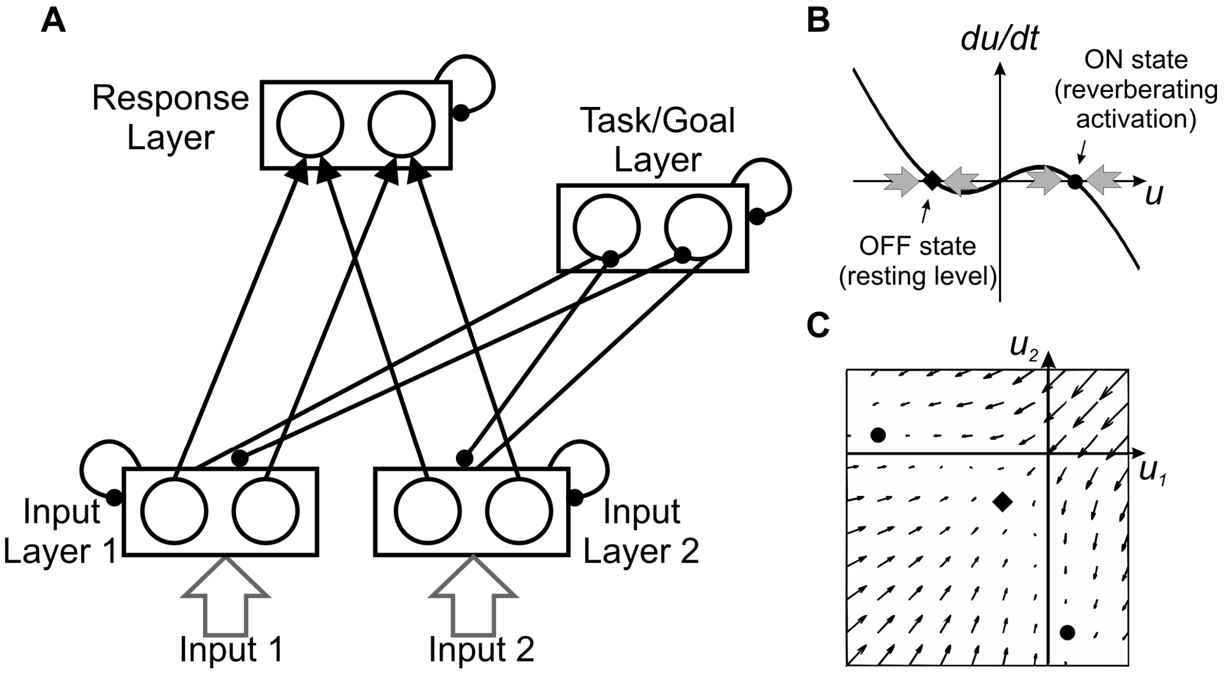


Figure 2

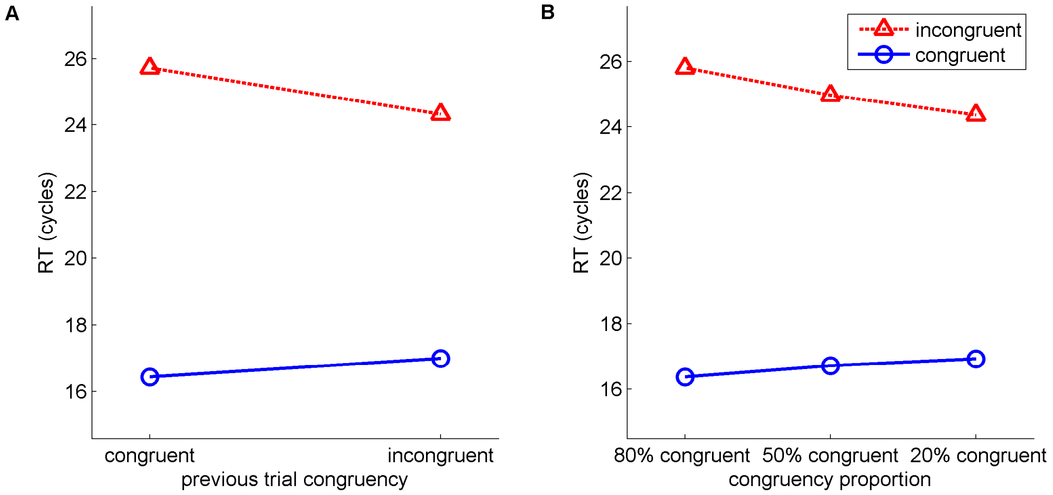

A
RT (cycles)
26
24
22
20
18
16
congruent
incongruent
previous trial congruency
B
RT (cycles)
26
24
22
20
18
16
80% congruent
50% congruent
20% congruent
congruency proportion
incongruent
congruent

Figure 3

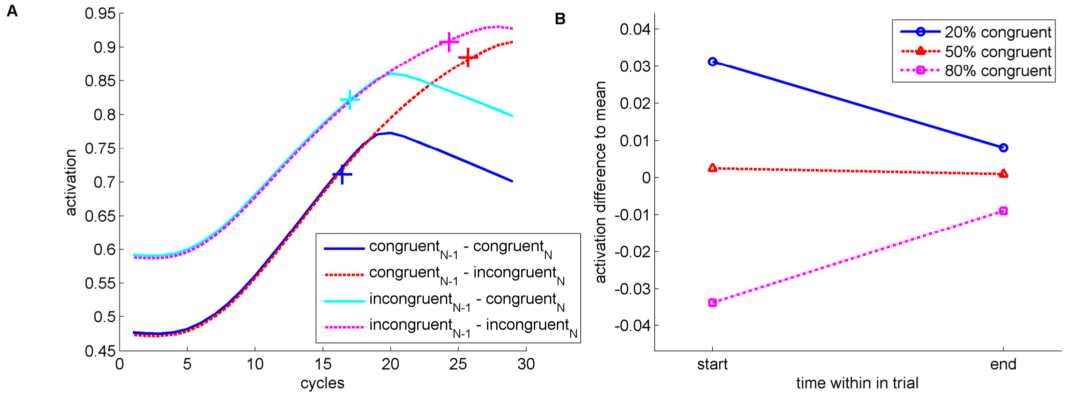
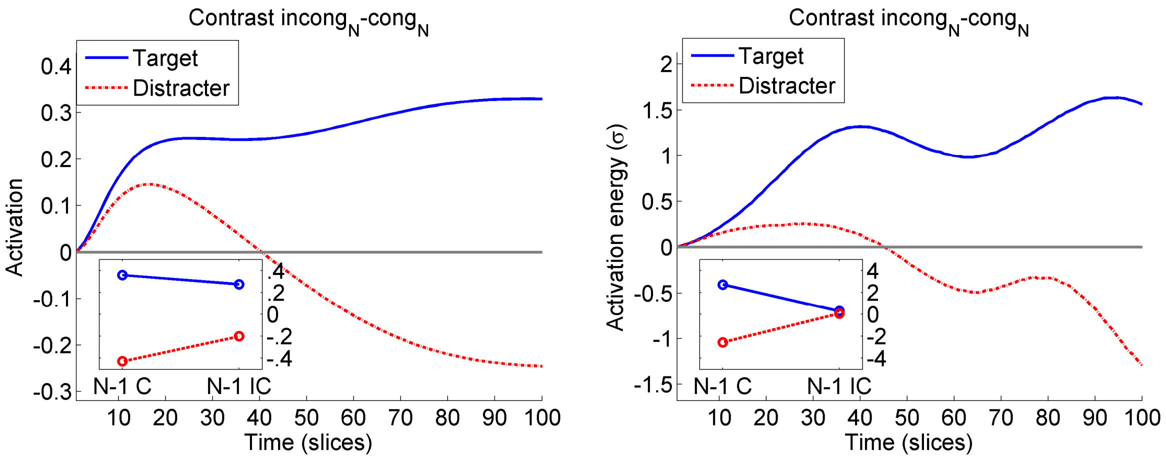

A
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
activation
0
5
10
15
20
25
30
cycles
congruent$_{N-1}$ - congruent$_N$
congruent$_{N-1}$ - incongruent$_N$
incongruent$_{N-1}$ - congruent$_N$
incongruent$_{N-1}$ - incongruent$_N$
B
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
activation difference to mean
start
end
time within in trial
20% congruent
50% congruent
80% congruent

Figure 4


Contrast incong$_N$-cong$_N$
Target
Distracter
Activation
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
.4
.2
0
-.2
-.4
N-1 C
N-1 IC
10 20 30 40 50 60 70 80 90 100
Time (slices)
Contrast incong$_N$-cong$_N$
Target
Distracter
Activation energy (σ)
2
1.5
1
0.5
0
-0.5
-1
-1.5
4
2
0
-2
-4
N-1 C
N-1 IC
10 20 30 40 50 60 70 80 90 100
Time (slices)

-

Figure 5

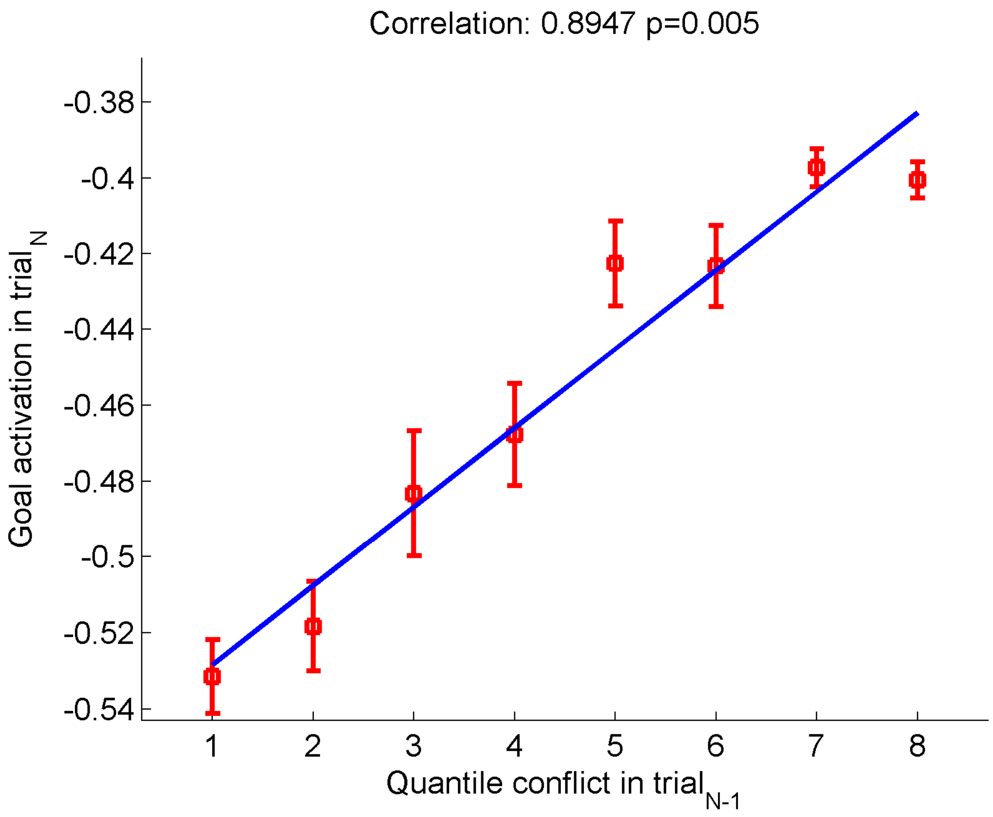

Correlation: 0.8947 p=0.005
-0.38
-0.4
-0.42
-0.44
-0.46
-0.48
-0.5
-0.52
-0.54
Goal activation in trial$_N$
1
2
3
4
5
6
7
8
Quantile conflict in trial$_{N-1}$

Figure Captions

Figure 1. Outline of the dynamic connectionist model. (A) Spatial structure of the model. There are two input layers, representing task relevant and task irrelevant information. The units in the response layer indicate the response. Units in the goal layer represent the currently active task. The input layers have feed forward connections to the response layer. Between the input and the goal layer are reciprocal excitatory or inhibitory connections. Input information excites its accordant goal and inhibits other goals, while goal activation excite according input units and inhibit other input, biasing processing in the input layer to task relevant information. (B) Activation dynamics of goal units. The x-axis shows the current activation of a unit (*u*), the y-axis shows the tendency to change for this current activation (*du/dt*). A stable activation state shows no tendency to change and hence must lie on the x-axis (*du/dt* = 0). Due to the sigmoid activation function, units exhibit two stable (attractor) states, with the neighboring points being attracted to these states (indicated by grey arrows). Hence, goal units exhibit a stable OFF state (marked by a diamond) and a stable ON state (marked by a circle), and each of these states is surrounded by their attractor basin. (C) Dynamics of the coupled system of goal units illustrated as a vector field. The current activation the two goal units is now mapped to one dimension each (*u1* and *u2*). The tendency to change is now indicated by arrows, with the length of each arrow indicating the strength of the tendency. Since, both goal units are coupled by inhibitory connections, the coupled system exhibits multistable dynamics with three possible stable states (both goals OFF, marked by the diamond; goal 1 ON, or goal 2 ON, marked by

circles). Within the respective attractor basins around these stable states, temporary modulations of activation are possible without changing the overall state of the goal system..

Figure 2. RT results of the simulation. Figure 2A shows the overall congruency and conflict adaptation effect, mirroring the results of Gratton and colleagues (1992, experiment 1, figure 1). Figure 2A illustrates the congruency proportion effect, mirroring the results of Gratton and colleagues (1992, experiment 2, figure 6)

Figure 3. Activation of the relevant goal unit in the goal layer. Figure 3A shows the average activation dynamics for different types of congruency in the previous (N-1) and the current (N) trial. Average response times are marked by a cross. Within a trial, the model moves from low activation of the goal unit to higher activation of the goal unit. At the start of trials following congruent trials, the goal unit is in a lower activation state than at the start of trials following incongruent trials. At the end of incongruent trials, the goal units is in a higher activation state than at the end of congruent trials. This difference opens up after the end of the shorter congruent trials, with incongruent trials leaving more time for the interaction dynamics to drive the goal units activation supporting the resolution of conflict. Figure 3B shows the activation at the start and the end of high conflict trials (incongruent trials following previously congruent trials): The data groups of high frequency congruent trials (80%), equal frequency congruent trials (50%) and low frequency congruent trials (20%) is subtracted from the data averaged across these groups. The difference at the start of trial indicates the different levels of preparedness for conflict under different frequencies

of conflict. While this difference decreases within a trial, the smaller gap at the end of the trial indicates a difference in adaptation to conflict after one conflict trial.

Figure 4. Activation contrast incongruent-congruent as a function of normalized time (100 time slices for every trial). Left panel: The contrast for the model's activated input units. Right panel: The contrast for the amplitude of frequency tagged EEG data from a continuous conflict adaptation study (Scherbaum et al., 2011). Simulation and real data show contrast enhancement between relevant target and irrelevant distracter information in conflict trials compared to no conflict trials. While simulation data show raw activation differences, EEG data was z-transformed to baseline data for aggregation of different tagging frequencies, hence both graphs showing different scales. The insets show the endpoints of the contrast enhancement (relative to the start at baseline) for trials following congruent or incongruent trials, indicating stronger contrast enhancement for previously congruent trials.

Figure 5. Mean activation of the task-correct goal unit in the current trial as a function of summed response unit activity on the previous trial (reflecting activation of lateral inhibition between conflicting responses). Following the procedure of Kerns and colleagues (2004), previous trials were divided into eight quantiles based on response conflict activity. For each quantile, the mean activity on the subsequent trials is plotted for the goal unit. Response conflict activity in the previous trial correlates with goal unit activity on the current trial despite no direct causal connection in the model.