A Neurocomputational Model of Automaticity and Maintenance of Abstract Rules

Sebastien Helie, F. Gregory Ashby

***Abstract*—Rule-guided behavior is essential in quickly adapting to one's ever-changing environment. In particular, learned rules can quickly be used in new contexts or applied to new stimuli (which confers an advantage over restricting learning to perceptual – motor associations). Here, we propose a new neurocomputational model of automaticity in rule-guided behavior. The proposed model assumes two parallel neural pathways corresponding to "naïve" and "expert" rule use. The development of automaticity is characterized by a transfer of control of rule-guided behavior from a pathway mediated by the prefrontal cortex to a direct parietal - premotor pathway. The model includes differential equations that describe voltage changes in the relevant brain areas and difference equations that describe the Hebbian learning. A simulation shows that the model accounts for some critical single-cell recording data from several key brain areas as well as some important behavioral results.**

I. Introduction

RULE-guided behavior is essential in adapting one's behavior to the ever changing environment [1]-[2]. In particular, rule-use allows for a direct generalization of performance to new stimuli or new situations [3]. Rule learning is to be contrasted with associative learning, where performance with new stimuli or in new situations is (or nearly is) at chance [4].

Early clinical studies have shown that damage to the prefrontal areas impairs performance in a rule-based card sorting task [5]-[6]. As a result, most neurobiological studies of rule-use and rule learning have focused on the role of the prefrontal cortex (e.g., [2], [7]-[8]; for a review, see [4]). However, more recent evidence suggests that other cortical areas, such as the premotor cortex, also play an important role in rule application [4], [9]-[10]. Moreover, it was shown that rule cells are present in the premotor cortex of expert monkeys [9], and that premotor rule cells fire earlier and more strongly than rule cells in the prefrontal cortex [10].

In a recent paper, Ashby and his colleagues [11] proposed that automaticity in non rule-guided categorization could be explained by a shift of behavioral control from a longer (subcortical) pathway to a shorter (cortical) pathway. Here, we propose a similar theory of rule-guided behavior, where automaticity in rule application can be explained by the shift from a longer cortical pathway mediated by the prefrontal cortex to a shorter cortical pathway feeding directly into the premotor cortex.

Manuscript received March 10, 2009. This work was supported in part by the U.S. Army Research Office through the Institute for Collaborative Biotechnologies under grant W911NF-07-1-0072.

S. Helie is with the Psychology Department, University of California, Santa Barbara, CA 93106-9660 USA (phone: 805-893-7909; fax: 805-893-4303; e-mail: helie@psych.ucsb.edu).

F. G. Ashby is with the Psychology Department, University of California, Santa Barbara, CA 93106-9660 USA (e-mail: ashby@psych.ucsb.edu).

II. Theory

The theory is illustrated by Fig. 1. As can be seen, the model is composed of two pathways: a long ("naïve") pathway representing "non-automatic" rule application (center column) and a short ("expert") pathway representing "automatic" rule application (right column). The left column (dashed box) is used for the maintenance of information in working memory and is directly borrowed from [12].

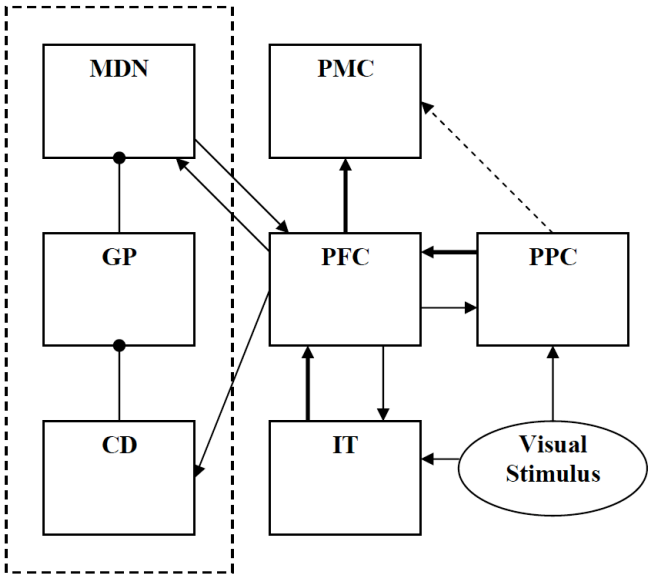

Fig. 1. Biological model of automaticity in rule-guided behavior. Arrows represent activation, circles represent inhibition, and dotted arrows represent learnable synapses (the line thickness represents the synapse strength). The left column (dashed box) is borrowed from FROST [12] and represents working memory, the middle column represents the "naïve" pathway, and the right column represents the "expert" pathway. MDN = Medial Dorsal Nucleus of the thalamus, GP = Globus Pallidus, CD = Caudate Nucleus, PMC = PreMotor Cortex, PFC = PreFrontal Cotex, IT = InferoTemporal Cortex, and PPC = Posterior Parietal Cortex.

A. The "Naïve" Pathway

In the naïve pathway, the visual stimulus is initially represented in the inferotemporal cortex (IT). Single cell recordings have shown that object-level representations are present in monkey IT [3], [10]. These stimulus cells in IT in turn activate stimulus representations in the prefrontal cortex (PFC), where information is represented at the categorical level in monkeys [3]. Feedback loops between stimulus cells in the PFC and stimulus cells in IT allow for a richer object representation in IT [3].

Rule cells that fire when a particular rule is applied, regardless of the stimulus or the source of rule activation, are also found in the PFC [4], [7], [9]-[10]. Rule cells are thought to bias synapses between the stimulus cells in the PFC and motor planning cells in the premotor cortex ([2]; more later).

The categorical representation of the stimulus in the PFC is connected to motor planning cells in the premotor cortex (PMC). Motor planning cells in the PMC fire when a particular response has to be initiated, regardless of stimulus or rule [4], [9]. Because each stimulus cell in the PFC can be associated to a different motor planning cell in the PMC depending on the applied rule, the synapses between the PFC stimulus cells and the PMC motor planning cells are all equally strong. However, the activated rule cell in the PFC inhibits the synapses between the PFC stimulus cells and the PMC motor planning cells that would lead to incorrect motor responses and excites the synapses between the PFC stimulus cells and the appropriate PMC motor planning cells. As such, a rule in the model is represented by a pattern of biasing synapses [2].

B. The "Expert" Pathway

Recent evidence suggests that the posterior parietal cortex (PPC) represents more than just spatial information [1], [13]. In particular, PPC cells have been shown to represent categorical level information [14] and are directly connected to the PMC [15]. Hence, in the model, the visual stimulus also activates stimulus cells in the PPC. These PPC stimulus cells are connected with the categorical stimulus cells in the PFC to form a closed loop [12]. This loop keeps the stimulus cells activated in the PPC even after the visual stimulus has disappeared [12], [16]. The categorical stimulus cells in the PPC are connected to the motor planning cells in the PMC.

Initially, the connections between the PPC stimulus cells and the PMC motor planning cells are weak (or inexistent). However, repeated simultaneous activation of the PPC stimulus cells (by lower-level visual areas) and the PMC motor planning cells (by the PFC cells in the "naïve" pathway) allows for synapse strengthening using Hebbian learning [11]. When the synapse between the PPC stimulus cells and the PMC motor planning cells are sufficiently strong, the "naïve" pathway is no longer required to produce the desired behavior and rule application is "automatic".

It should be noted that, because all the stimulus cells in the PPC can be associated with all of the motor planning cells in the PMC (depending on which rule is applied), all synapses in the "expert" pathway are, on average, equally strengthened. Hence, a rule cell is required to inhibit synapses that would lead to incorrect motor responses and excite synapses that would lead to correct motor responses (as in the "naïve" pathway). Rule cells displaying the same characteristics as the PFC rule cells have been found in the PMC of expert monkeys [4], [9]. These rule cells in the PMC play the same biasing role as their counterparts in the PFC [10]. The main distinction between PFC and PMC rule cells is that rule cells in the PMC of expert monkeys fire earlier and more strongly than rule cells in the PFC [10], thus supporting the transfer of control predicted by the model.

C. Working Memory

Working memory in the model is borrowed from [12]. In FROST, one of the roles of the PFC is to activate the caudate nucleus (CD) when a stimulus to be remembered becomes unavailable. Firing cells in the CD inhibit cells in the globus pallidus (GP), which by default were tonically firing to inhibit the medial dorsal nucleus of the thalamus (MDN). After the GP cells have stopped firing, the MDN cells start firing in a loop with the stimulus cells in the PFC, which maintains the PFC categorical representations activated after activation from the IT and the PPC stimulus cells has stopped. This subcortical pathway has successfully been used to simulate many working memory phenomena at the cellular and behavioral levels [12].

III. Computational Model

The Izikevich simple two-dimensional model was used to implement the cortical cells [17]. In particular, all the cortical cells were regular spiking neurons [17]. The general equation is:

$$100\frac{d\left(v_{a(i)}\right)}{dt} = 0.7\left(v_{a(i)} + 60\right)\left(v_{a(i)} + 40\right) - u_{a(i)} + I_{a(i)} + \sigma \qquad (1)$$

$$\frac{d\left(u_{a(i)}\right)}{dt} = 0.03 \times \left[-2\left(v_{a(i)} + 60\right) - u_{a(i)}\right]$$

where $v_{a(i)}$ represents the membrane potential of neuron i in area a, $u_{a(i)}$ represents the recovery current of neuron i in area a, $I_{a(i)}$ is the input to neuron i in area a, and σ represents zero-mean Gaussian noise (where σ is the standard deviation). If (1) reaches 35mV, the membrane potential is reset to -50mV and the recovery current is increased by 100mV (i.e., if $v_{a(i)} > 35$, $v_{a(i)} = -50$ and $u_{a(i)} = u_{a(i)} + 100$).

The constants in (1) where estimated by fitting the equation to single cell recordings in patch clamp experiments [17]. The choice of a single cortical cell model is consistent with neurobiological results (for a review, see [17]) and reduces the model complexity.

The caudate nucleus cell is modeled by a (different) two-dimensional simple model, while the globus pallidus and thalamus cells were modeled with one-dimensional quadratic integrate-and-fire models [17] (see *Section III.E*). Simpler modeling of the subcortical cells was motivated by the secondary role played by working memory in the present application.

The postsynaptic effects of a spike generated in a presynaptic cell were modeled via the α-function: Every time the presynaptic cell spiked, the following input was delivered to the postsynaptic cell [18]:

$$\alpha(t) = \frac{t}{\lambda} e^{\frac{\lambda - t}{\lambda}} \qquad (2)$$

where λ is a free parameter representing the time lag. The values given to all the free parameters are defined in Table I.

A. Inferotemporal Cortex

The change in activation in the IT stimulus cells is described by (1), with the following input:

$$I_{IT(i)} = I_{stim(i)} + w_{PFC(i)\to IT(i)} \times \alpha\left(v_{PFC,stim(i)}\right) \tag{3}$$

where $I_{stim(i)}$ represents the input from lower-level visual areas (here, a square activation function; see Table I), $w_{PFC(i)\to IT(i)}$ represents the strength of the synapse between PFC stimulus cell i and IT stimulus cell i, and $v_{PFC,stim(i)}$ is the input from PFC stimulus cell i.

B. The Prefrontal cortex

There are two types of modeled cells in the prefrontal cortex: stimulus cells [2]-[3] and rule cells [8]-[9]. Like the stimulus cell in IT, the stimulus cell in the PFC is modeled using (1). However, the input is different:

$$\begin{aligned} I_{PFC,Stim(i)} = w_{IT(i)\to PFC(i)} \times \alpha\left(v_{IT}\right) + w_{MDN\to PFC(i)} \times \alpha\left(v_{MDN}\right) \\ + w_{PPC(i)\to PFC(i)} \times \alpha\left(v_{PPC(i)}\right) \end{aligned} \tag{4}$$

where $w_{IT(i)\to PFC(i)}$, $w_{MDN\to PFC(i)}$, $w_{PPC(i)\to PFC(i)}$ represent synapse strengths from IT, MDN, and PPC respectively, and $v_{IT(i)}$, v_{MDN}, and $v_{PPC(i)}$ represent inputs from IT, the MDN, and the PPC respectively.

The second type of cells modeled in the PFC is rule cells. PFC rule cells are cortical and modeled by (1). However, unlike the previously modeled stimulus cells, incoming activation to the PFC rule cells comes from outside the model. The value given to $I_{PFC,Rule(i)}$ is shown in Table I.

C. Posterior Parietal Cortex

To reduce the number of assumptions/parameters, stimulus cells in the PPC were modeled in exactly the same way as the stimulus cells in IT. As such, the transmission is described by (1) and the input is described by:

$$I_{PPC(i)} = I_{stim(i)} + w_{PFC(i)\to PPC(i)} \times \alpha\left(v_{PFC,stim(i)}\right) \tag{5}$$

where where $I_{stim(i)}$ represents the input from lower-level visual areas (the same as in (3); see Table I), $w_{PFC(i)\to PPC(i)}$ is the synapse strength between the stimulus cells in the PFC and the PPC, and $v_{PFC,stim(i)}$ represents the input from PFC stimulus cell i.

D. Premotor Cortex

There are two types of modeled cells in the premotor cortex. The first type is the rule cell [4], [9]-[10]. Because the rule cells in the premotor cortex are thought to play a role similar to the rule cells in the PFC [10], they are modeled using the same equation (1), but with a different input. The input, $I_{PMC,Rule(i)}$ comes from outside the model and is defined in Table I.

The second type of premotor cortex cells included in the model is motor planning cells [4], [9]-[10]. These cells represent the output of the model (primary motor areas are outside the scope of the present work). The motor planning cells are described by (1) with the following input:

$$\begin{aligned} I_{PMC,motor(i)} = \sum_k w_{PFC,stim(k)\to PMC(i)} \times \alpha\left(v_{PFC,stim(k)}\right) \\ + \sum_k w_{PPC(k)\to PMC(i)} \times \alpha\left(v_{PPC(k)}\right) \\ + w_{PFC,rule(activated)\to PMC(i)} \times \alpha\left(v_{PFC,rule(k)}\right) \\ + w_{PMC,rule(activated)\to PMC(i)} \times \alpha\left(v_{PMC,rule(k)}\right) \end{aligned} \tag{6}$$

where $w_{PFC,stim(k)\to PMC(i)}$, $w_{PPC(k)\to PMC(i)}$, $w_{PFC,rule(activated)\to PMC(i)}$, $w_{PMC,rule(activated)\to PMC(i)}$ are the synapse strengths between motor planning cell i in the PMC and stimulus cell k in the PFC, stimulus cell k in the PPC, the activated rule cell in the PFC, and the activated rule cell in the PMC respectively. $v_{PFC,Stim(k)}$, $v_{PPC(k)}$, $v_{PFC,Rule(k)}$, and $v_{PMC,Rule(k)}$ represent the input from stimulus cells k in the PFC and the PPC and from rule cells k in the PFC and the PMC respectively.

In (6), the first two lines represent stimulus activation from the "naïve" and "expert" pathways (respectively), while the third and fourth lines represent the bias from rule cells in the "naïve" and "expert" pathways (respectively). If motor planning cell i is the appropriate response when a particular combination of rule and stimulus is presented, $w_{PFC,rule(activated)\to PMC(i)}$ and $w_{PMC(activated)\to PMC(i)}$ are positive; otherwise, they are negative (see Table I).

E. Subcortical areas

A few subcortical areas have been included in the model to implement working memory (see the dashed box in Fig. 1). This subcortical network was borrowed from FROST [12], an already established biological model of working memory. However, the model was re-implemented using one- and two-dimensional model cells from [17]. Because only one item needed to be kept in working memory in the simulation included in this paper (see *Section IV*), only one cell from each subcortical area was included and subscripts were dropped for clarity. However, the model can be directly generalized to cases with more than one cells [12]. Also, note that noise was not included in the subcortical cell models, as single cell recording data were not modeled here (but see [12]).

The first cell included in the dashed box of Fig. 1 is a CD cell. This cell receives activation from the PFC when the visual stimulus disappears and is described by:

$$\begin{aligned} 50\frac{d\left(v_{CD}\right)}{dt} = \left(v_{CD}+80\right)\left(v_{CD}+25\right) - u_{CD} + I_{CD} \\ 100\frac{d\left(u_{CD}\right)}{dt} = \left[-20\left(v_{CD}+80\right) - u_{CD}\right] \end{aligned} \tag{7}$$

where v_{CD} represents the membrane potential, u_{CD} represents the recovery current, and I_{CD} is the input from the PFC (see Table I). If (7) reaches 40mV, the cell membrane is reset to -55mV and the recovery current is increased by 150 (i.e., if

$v_{CD} > 40$, $v_{CD} = -55$ and $u_{CD} = u_{CD} + 150$). Like in (1), the constants were estimated by fitting single cell recordings in patch clamp experiments [17].

The GP is modeled using a one-dimensional quadratic integrate-and-fire model [17]. This cell tonically fires at rest and stops firing when it receives activation from the CD. It is described by:

$$\frac{d(v_{GP})}{dt} = 2[0.7(v_{GP} + 60)(v_{GP} + 40) + 71 - \alpha(v_{CD}) \times w_{CD \to GP}] \quad (8)$$

where v_{GP} represents the membrane potential and $w_{CD \to GP}$ represents the synapse strength between the CD and GP cells. If (8) reaches 35mV, the cell membrane is reset to -50mV (i.e., if $v_{GP} > 35$, $v_{GP} = -50$). The constants were estimated by fitting single cell recordings in patch clamp experiments [17].

When the GP cell stops firing, the MDN cell can start firing, thus initiating a loop to keep the stimulus cells active in the PFC. The MDN cell is also modeled using a one-dimensional quadratic integrate-and-fire model [17]:

$$\frac{d(v_{MDN})}{dt} = 0.7(v_{MDN} + 60)(v_{MDN} + 40) + 71 - \alpha(v_{GP}) \times w_{GP \to MDN} + \alpha(v_{PFC,stim(i)}) \times w_{PFC \to MDN} \quad (9)$$

where v_{MDN} represents the membrane potential, $v_{PFC,stim(i)}$ is the input from PFC stimulus cell i, $w_{GP \to MDN}$ represents the synapse strength between the GP and the MDN cells, and $w_{PFC \to MDN}$ represents the synapse strength between the PFC stimulus cell and the MDN cell. The threshold and peak value are the same as in (8). The constants were estimated by fitting single cell recordings in patch clamp experiments [17].

F. Learning

The only modifiable synapses in the model are those connecting the PPC stimulus cells to the PMC motor planning cells ($w_{PPC(k) \to PMC(i)}$). These cortical connections are learned using Hebbian learning [11]:

$$w_{PPC(k) \to PMC(i)}(t+1) = w_{PPC(k) \to PMC(i)}(t) + \beta \times \Delta w_{PPC(k) \to PMC(i)}(t) \quad (10)$$

$$\Delta w_{PPC(k) \to PMC(i)}(t) = f\left[\sum_t I_{PMC,motor}(t)\right]^+ f\left[\sum_t g(v_{PMC,motor}(t)) - \theta\right]^+ - f\left[\sum_t I_{PMC,motor}(t)\right]^+ f\left[\theta - \sum_t g(v_{PMC,motor}(t))\right]^+$$

where β is a scaling parameters, $v_{PMC,motor}(t)$ is the output of the PMC motor planning cell at time t, $I_{PMC,motor}(t)$ is the input of the PMC motor planning cell at time t (6), $f[\bullet]^+$ is a clipping function that equals zero if the operand is negative (and otherwise equals the operand), $g(\bullet)$ scales the operand between 0 and 1, and θ is the activation threshold of the NMDA receptors. The first term (positive) represents long-term potentiation while the last term (negative) represents long-term depression.

IV. Simulation

The new model was used to simulate data from three adult rhesus monkeys in a "Same" – "Different" task [9]-[10]. All the free parameters were set as in Table I throughout.

A. Task Description

Each trial was composed of four different events. The first event was the simultaneous presentation of the first stimulus (i.e., the "sample") and a cue indicating what rule to apply in the trial. The monkeys were trained on two different rules: the "match" rule and the "nonmatch" rule (more later). The first event lasted 800ms. The second event was a delay during which no stimulus was presented (duration: 1500ms). The third event was marked by the presentation of a second stimulus (i.e., the "test") for a 500ms period. Afterward, the test stimulus disappeared and the monkey had 500ms to provide a response. If the rule was "match" and the sample and test stimuli were the same, the monkey needed to *release* a lever. If the rule was "match" and the sample and test stimuli were different, the monkey had to *hold* the lever.

The "nonmatch" rule was the opposite of the "match" rule. Hence, the monkey had to *release* the lever if the sample and test stimuli were different and *hold* the lever if they were the same. The timing of a trial is shown in Fig. 2.

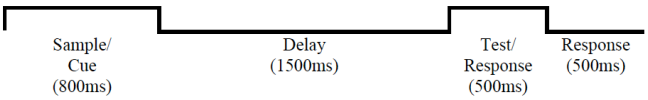

Fig. 2. Event timing of a trial in the "Same"-"Different" task. Up states correspond to stimulus presentations while down states correspond to blank screens. The lengths of the events are scaled.

B. Simulation Setup

The simulation was made to reproduce as closely as possible the experimental controls in [9]-[10]. The timing of the events was identical, and all four types of trials were simulated (i.e., "match" – same, "match" – different, "nonmatch" – same, "nonmatch" – different). The model included two stimulus cells in IT, two stimulus cells in the PFC, two rule cells in the PFC, two stimulus cells in the PPC, two rule cells in the PMC, two motor planning cells in the PMC ("hold", "release"), and one cell in each of the subcortical areas (CD, GP, and MDN). The stimulus cells in IT, the PFC, and the PPC all represented the same set of two objects. The rule cells in the PFC and PMC both represented the same rule set ("match", "nonmatch").

Each stimulus cell in IT was only connected to the stimulus cell representing the same object in the PFC. Similarly, each stimulus cell in the PPC was only connected to the stimulus cell representing the same object in the PFC. The stimulus cells in the PFC were connected to both motor planning cells in the PMC, and to the MDN cell.

A different PFC area (not included in the model) was connected to the CD. The CD cell was connected to the GP cell, and the GP cell was connected to the MDN cell. The

MDN cell was connected to the stimulus cells in the PFC. This closed loop represented working memory [12].

TABLE I
VALUES ASSIGNED TO THE FREE PARAMETERS

Parameter	Equations	Value
λ	(2)	60
σ	(1)	4000
$I_{stim(i)}$	(3), (5)	{0, 5000}[a]
$I_{PFC,rule(i)}$	-	{0, 5000, 10000}[b]
$I_{PMC,rule(i)}$	-	{0, -4.93t + 12496, 7500}[c]
I_{CD}	(7)	{0, 15000}[d]
$w_{PFC(i)\rightarrow IT(i)}$	(3)	25
$w_{IT(i)\rightarrow PFC(i)}$	(4)	60
$w_{MDN\rightarrow PFC(i)}$	(4)	15
$w_{PPC(i)\rightarrow PFC(i)}$	(4)	55
$w_{PFC(i)\rightarrow PPC(i)}$	(5)	25
$w_{PFC,stim(i)\rightarrow PMC(i)}$	(6)	65
$w_{PFC,rule(i)\rightarrow PMC(i)}$	(6)	±20
$w_{PMC,rule(i)\rightarrow PMC(i)}$	(6)	±25
$w_{CD\rightarrow GP}$	(8)	0.01
$w_{PFC\rightarrow MDN}$	(9)	0.25
$w_{GP\rightarrow MDN}$	(9)	0.01
β	(10)	2×10^{-10}
$\boldsymbol{\theta}$	(10)	400

[a]The input to the stimulus cells in IT and the PPC was 5000 during stimulus presentations and 0 in between stimulus presentations (see Fig. 2).

[b]The input to the activated rule cell in the PFC was 0 for the first 400ms, 5000 for the following 2100ms, and 10000 for the rest of the trial. These values were chosen to reproduce single cell recordings in [10].

[c]The input to the activated rule cell in the PMC was 0 for the first 100ms, a linear function of time (in ms) for the following 2200ms, and 7500 for the rest of the trial. These values were chosen to reproduce single cell recordings in [9].

[d]The input to the CD cell was 0 during stimulus presentations and 15000 in between stimulus presentations (see Fig. 2).

The rule cells in the PFC and the PMC were connected to the motor planning cells in the PMC. If the rule was "match" and the sample and test stimuli were the same, the "hold" motor planning cell was inhibited in the PMC and the "release" cell was excited. If the rule was "match" and the sample and test stimuli were different, the "release" motor planning cell was inhibited in the PMC and the "hold" cell was excited. The "nonmatch" rule cell produced the opposite pattern of motor planning cell bias in the PMC.

Finally, the connections between the stimulus cells in the PPC and the motor planning cells in the PMC were initially inexistent ($w_{PPC(k)\rightarrow PMC(k)}(0) = 0$). These connections were learned using (10).

C. Simulating a trial

A trial was initiated by the simultaneous activation of a rule cell in the PFC and the PMC ("match" or "nonmatch") and a stimulus cell in IT and the PPC (object #1 or object #2). These stimulus cells were used to activate the corresponding stimulus cell in the PFC.

The beginning of the second event was marked by a release of the activation of the stimulus cells in IT and the PPC and the activation of the CD cell by the PFC. Activation of the CD cell initiated the working memory loop, which kept the stimulus cells activated in the PFC, (and in the PPC and IT; to a lesser extent).

The third event started with the activation of a stimulus cell in IT and the PPC (object #1 or object #2). Like in the first event, the activated stimulus cells in IT and the PPC activated the corresponding stimulus cell in the PFC. If the test stimulus (from the third event) was the same as the sample stimulus (from the first event), the PFC and PPC stimulus cells representing the object fired more strongly, as the new activation from the third event was added to the working memory activation (and the stimulus cell representing the other object fired weakly because neither stimulus/event activated it). In contrast, if the test and sample stimuli were different, both stimulus cells fired moderately, because one of the cells was still activated by working memory (corresponding to the sample stimulus) and the other was newly activated by the test stimulus. This difference in firing rate allowed distinguishing same trials from different trials.

Finally the fourth event was marked by the activation of the motor planning cells in the PMC by the stimulus cells in the PFC and the PPC. The contribution from the PPC stimulus cells was negligible at first, but eventually drove the motor planning cells firing rate when rule application became "automatic" (because Hebbian learning strengthened the synapses between the motor planning cells in the PMC and the stimulus cells in the PPC). In addition to motor planning cell activation, the fourth event was marked by biasing activation sent by the rule cells in the PFC and the PMC.

D. Results

The aim of this simulation was twofold. First, the neurocomputational model should be able to capture neurobiological properties (e.g., single-cell recordings). Second, the behavior of the model should display a switch of control from the "naïve" pathway to the "expert" pathway after training. These two goals are now discussed.

1) Neurobiological Properties

To evaluate the propensity of the model to reproduce detailed neurobiological properties, the firing rate of the simulated cells was compared to monkey single cell recordings in the "Same" – "Different" task [9]-[10]. The left panel of Fig. 3 shows the firing rate of a simulated cell in IT. As can be seen, the firing pattern is qualitatively similar to the firing pattern of a monkey IT picture selective cell in a "Same" – "Different" task (right panel of Fig 3; see [10]). Both the simulated and monkey neurons fired if and only if a particular sample stimulus was presented.

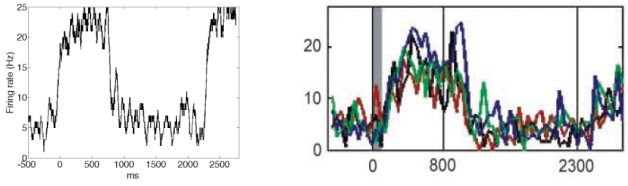

Fig. 3. Picture selective neuron in IT. Left: Simulated cell activation in IT by the model. Right: Single cell recording from a picture selective cell in IT [10]. The color represents the rule cue (not modeled here). In each panel, the x-axis is zeroed at the onset of the sample stimulus.

Stimuli are also represented in the PFC during a "Same" – "Different" task [4], [9]. The left panel of Fig. 4 shows the simulated firing rate of a PFC stimulus cell. As can be seen, the complex firing pattern of real PFC picture selective neurons is well matched by the new model (see the right panel of Fig. 4; [9]).

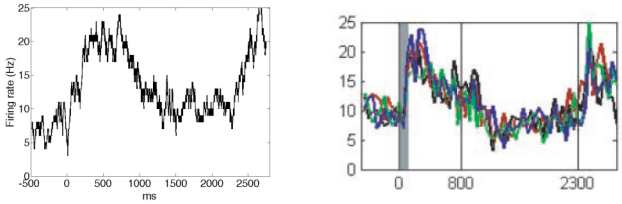

Fig. 4. Picture-selective neurons in the PFC. Left: Simulated cell activation of a stimulus cell in the PFC by the model. Right: Single cell recording of a picture selective cell in the PFC [9]. The color represents the rule cue (not modeled here). In each panel, the *x*-axis is zeroed at the onset of the sample stimulus.

Overall, the new model was successful at reproducing cell firing patterns from monkey single cell recordings [9]-[10] with a fairly simple cell model [17]. The main difference between the simulated cells and the monkey cells is located at the onset of activity (slightly faster in IT; slightly slower in the PFC). This small quantitative difference is unlikely to be a shortcoming of the proposed model architecture but a consequence of the simplicity of the cell model used. However, the added complexity of a better fitting cell model would not further the understanding of automaticity in rule-guided behavior, which was the focus of the present modeling work. Thus, the cell model proposed in [17] provided a useful compromise.

2) Behavior Control

After 20 trials of practice, the synapse strengths between the PPC stimulus cells and the PMC motor planning cells were $w_{PPC \rightarrow PMC}(20) \approx \{84, 84; 91, 91\}$. Fig. 5 shows the resulting simulated PMC motor planning cell activation. As can be seen in the top panel, the behavior was initially controlled by the PFC (dark gray line), but the synapses in the "expert" pathway eventually became stronger (bottom panel) and took control of the motor planning cell responses in the PMC (light gray line). The bottom panel of Fig. 5 also shows that when rule-guided behavior is automatic, the cell activation increases sooner (~150 ms after stimulus onset, compared with ~250 ms after stimulus onset in the top panel). This faster increase of activation was caused by the stronger connections in the shorter ("expert") pathway after training. Hence, the new model displays the learning capacity necessary to produce "automaticity" in rule application by switching neural pathway to avoid relying on rule application in the PFC.

V. Conclusion

This paper introduced a new neurobiologically detailed model of automaticity in rule-guided behavior. In the new framework, automaticity was defined by rule application without the involvement of the prefrontal cortex. More precisely, it was shown how a longer pathway mediated by the prefrontal cortex could train cortical connections between sensory areas and premotor areas. Learning was done by Hebbian learning [11].

The proposed model shares some similarities with the SPEED model of automaticity [11]. In both cases, a slower, longer pathway is used to train a shorter cortical pathway using Hebbian learning. The main difference between the new model and SPEED is that both pathways are cortical in the new model. This is in correspondence with rule application being considered a higher-level cognitive process than associative/procedural learning [5]. Hence, this work can be considered complementary to the SPEED model, which has been used to simulate automaticity in non-rule-guided behavior. The new model proposes a new definition of automaticity in rule-guided behavior, i.e., the bypassing of the prefrontal cortex in favor of a shorter, more direct neural circuit. Future work should be devoted to applying the new model to human fMRI data and comparing automaticity in rule-based and non-rule-based behavior to determine if a common brain pathway is used. Also, more complex rule-guided behavior should be simulated (e.g., involving a bigger network).

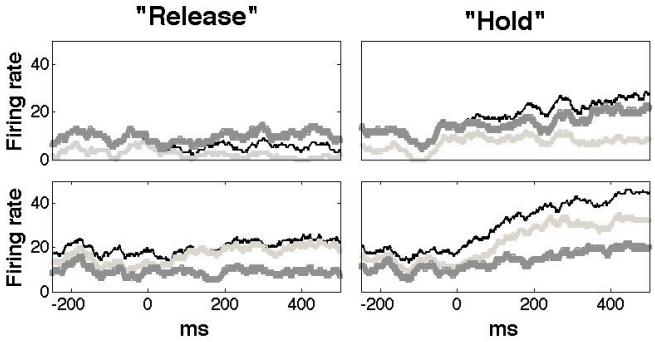

Fig. 5. Simulated activation from motor planning cells in the PMC on a trial where the rule was "nonmatch" and the stimuli were the same. The top panel shows the cells activations before training while the bottom panel shows the cells activations after 20 trials of training. The black line represents overall activation (PMC + PFC), the dark gray line represents PFC activation alone (from the longer pathway), while the light gray line represents activation from the shorter pathway (PMC only). The *x*-axis is zeroed at the onset of the test stimulus. The y-axis is scaled in hertz.

References

[1] S. A. Bunge, and M. J. Souza, "Neural representations used to specify action," in *Neuroscience of Rule-Guided Behavior*, S. A. Bunge, and J. D. Wallis, Eds. New York: Oxford University Press, 2007, pp. 45-65.

[2] E. K. Miller, and J. D. Cohen, "An integrative theory of prefrontal cortex function," *Anuual Review of Neuroscience*, vol. 24, pp. 167-202, 2001.

[3] D. J. Freedman, "Exploring the roles of the frontal, temporal, and parietal lobes in visual categorization," in *Neuroscience of Rule-Guided Behavior*, S. A. Bunge, and J. D. Wallis, Eds. New York: Oxford University Press, 2007, pp. 391-418.

[4] J. D. Wallis, "Single neuron activity underlying behavior-guiding rules," in *Neuroscience of Rule-Guided Behavior*, S. A. Bunge, and J. D. Wallis, Eds. New York: Oxford University Press, 2007, pp. 23-44.

[5] J. M. Fuster, "The prefrontal cortex – An update: Time is of the essence," *Neuron*, vol. 30, pp. 319-333, 2001.

[6] S. A. Bunge, "How we use rules to select actions: A review of evidence from cognitive neuroscience," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 4, pp. 564-579, 2004.

[7] W. F. Asaad, G. Rainer, and E. K. Miller, "Neural activity in the primate prefrontal cortex during associative learning," *Neuron*, vol. 21, pp. 1399-1407, 1998.
[8] I. M. White, and S. P. Wise, "Rule-dependent neuronal activity in the prefrontal cortex," *Experimental Brain Research*, vol. 126, pp. 315-335, 1999.
[9] J. D. Wallis, and E. K. Miller, "From rule to response: Neuronal processes in the premotor and prefrontal cortex," *Journal of Neurophysiology*, vol. 90, pp. 1790-1806, 2003.
[10] R. Muhammad, J. D. Wallis, and E. K. Miller, "A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum," *Journal of Cognitive Neuroscience*, vol. 18, pp. 974-989, 2006.
[11] F. G. Ashby, J. M. Ennis, and B. J. Spiering, "A neurobiological theory of automaticity in perceptual categorization," *Psychological Review*, vol. 114, pp. 632-656, 2007.
[12] F. G. Ashby, S. W. Ell, V. Valentin, and M. B. Casale, "FROST: A distributed neurocomputational model of working memory maintenance," *Journal of Cognitive Neuroscience*, vol. 17, pp. 1728-1743, 2005.
[13] D. J. Freedman, "Posterior parietal cortex: Space…and beyond," *Neuron*, vol. 42, pp. 881-883, 2004.
[14] L. J. Toth, and J. A. Assad, "Dynamic coding of behaviorally relevant stimuli in parietal cortex," *Nature*, vol. 415, pp. 165-168, 2002.
[15] S. P. Wise, D. Boussaoud, P. B. Johnson, and R. Caminiti, "Premotor and parietal cortex: Corticocortical connectivity and combinatorial computations," *Annual Review of Neuroscience*, vol. 20, pp. 25-42, 1997.
[16] M. Osaka, and N. Osaka, "Neural bases of focusing attention in working memory: An fMRI study based on individual differences," in *The Cognitive Neuroscience of Working Memory*, N. Osaka, R. H. Logie, and M. D'Esposito, Eds. New York: Oxford University Press, 2007, pp. 99-117.
[17] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press, 2007.
[18] W. Rall, "Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input, " *Journal of Neurophysiology*, vol. 30, pp. 1138-1168, 1967.