INNER SIMULATION SUSTAINING THE DELIBERATIVE PROCESS IN A COGNITIVE ARCHITECTURE

Othalia Larue[(a)], Pierre Poirier[(b)], Roger Nkambou[(a)]

[(a)] Computer Science department, GDAC Research Laboratory
[(b)] Philosophy Department
Université du Québec à Montréal
P.O. Box 8888Centre Ville Station, Montréal, Québec, Canada

[(a)] larue.othalia@courrier.uqam.ca, [(b)] poirier.pierre@uqam.ca, [(c)] nkambou.roger@uqam.ca

ABSTRACT

We propose a three level cognitive architecture for the simulation of cognitive phenomena. This architecture is based on Stanovich's tripartite framework (2010), a unified model of cognition, which provides an explanation of how deliberative (characterized by sequentiality) and adaptive (characterized by reactivity) human behaviour emerges from the interaction of three distinct cognitive levels (autonomous/reactive, algorithmic/cognitive control, and reflective). In previous work (Larue et al, 2012) we focused on the interaction of algorithmic and reactive level on a task evaluating cognitive control. In this paper, we focus on the interaction between reflective and algorithmic level. More precisely, thanks to a Wisconsin card Sorting task, a task that evaluates cognitive flexibility (the ability to change strategies), we study how cognitive decoupling (or inner simulation) supports the deliberative behaviour and hypothesis testing.

Keywords: Wisconsin Card sorting task, cognitive simulations, cognitive architecture, dual process theories

1. INTRODUCTION

Many cognitive architectures are designed with the intent of reproducing the human cognitive architecture, the human mind, either by replicating known fine-grained structures of the brain (Eliasmith, 2005) or by building systems with capacities that are functionally equivalent to those of humans. Designers of such architecture all face what might be called "the duality challenge." Evidence from many fields in cognitive psychology (psychology of reasoning, moral psychology, social psychology, etc.; see Evans 2008 for a review) and from cognitive neuroscience (see e.g., Goel, 2009) suggests that evolution may have built functionally incompatible (Sherry & Schacter, 1987) features into the human mind. On the one hand the mind is dynamical and reactive, simultaneously responding to many features in the environment in a seamless dynamical agent environment loop, while, on the other hand, it is sequential and rule following, applying explicitly learned rules one by one to plan its long-term behaviour or solve other complex problems.

One important source of evidence that helped create and sustain the duality challenge (at least in its contemporary version) is to be found in the various tests used by cognitive psychologists and neuropsychologists to assess cognitive performance: the Stroop task (attention), the Wisconsin card sorting test (WCST) (cognitive flexibility), the Wason selection task (logical reasoning), the Iowa Gambling task (decision making), and others. One of the early proposals was Evans' (1984) attempt to explain the bias observed in the Wason selection task (a task frequently used in the psychology of reasoning) by positing a competition between opposed heuristic (fast and automatic) and analytic (slow) processes. Explanations of the Stroop effect (Stroop, 1935) likewise often posit competing automatic processes, where attention, an opposed voluntary process, has to favor the less automatic of the two competing automatic processes. Similarly, many explanations of the perseverative errors in the Wisconsin Card Sorting Test rest on interactions between automatic and attentive processes. To address the duality challenge, the current trend in cognitive science and neuroscience is to posit dual system (Kahneman, 2011) or dual-process (Evans, 2008; Stanovich, 2010; Stanovich, 2011) theories, a "two-minds mind" (Frankish & Evans 2009), often called "System 1" and "System 2," where processes with features taken from a list of duals (implicit vs. explicit, automatic vs. voluntary, etc.) compete or collaborate to explain observed human behaviour. Designers of cognitive architectures in general have met the duality challenge by building hybrid architectures, interfacing dynamical and parallel systems such as neural networks with sequential and rule following systems, such as production systems. Dual-process theories (Evans, 2006; Sloman, 1996; Kahneman, 2011), however, have attracted much criticism in the literature (e.g., Keren &

Schul, 2009) where they are (rightly) said to be oversimplifications. One pressing problem is to account for the interaction between the opposing processes: how, for instance, can a voluntary process interrupt an automatic one; or again how can the output of a domain-specific process affect domain-general decision making. Call this the "interface problem" (of current-trend solutions to the duality challenge). Answers to the interface problem range from mere hand-waving (simply declaring that processes (somehow) interact) to nihilism (simply declaring that one set of processes do not exists – e.g., the mind is all (massively) modular) to despair (Fodor, 2000). Designers of cognitive architectures in general have dealt with the interface problem by converting the cognitive dichotomy into a paradigm (connectionist-symbolic) dichotomy (see for instance CLARION (Connectionist Learning with Adaptive Rule Induction ON line, Sun, 2004).

A few philosophers (e.g., Carruthers, 2006), computer scientists (e.g. Sloman & Chrisley, 2005) and psychologists, however, have addressed the interface problem directly, attempting to provide positive accounts of the means of interaction between mind's two minds. One such attempt is Stanovich's Tripartite Framework, which, paradoxically, begins by positing a three-minds mind: an Autonomous Mind, an Algorithmic Mind and a Reflective Mind. We use this model as the basis for the design of our architecture.

Our aim is to motivate the resulting architecture as a simulation tool. In previous work, we illustrated and validated the performance of the system on the two lower levels (autonomous and algorithmic minds), thereby demonstrating the adaptive behaviour of our system, by means of two variants of the Stroop task (classical and semantic) (Larue et al., 2012). In this paper, we focus on the interaction between the reflective mind and the algorithmic mind of the system. To do so, we added to our architecture a function that has been deemed, in Stanovich's tripartite framework, a key aspect in the production of deliberative behaviour: cognitive decoupling. We use a Wisconsin card sorting, a task which requires cognitive decoupling and cognitive flexibility to be performed, as a way to illustrate the system's performance.

2. RELATED WORK

2.1. Stanovich's tripartite model

2.1.1. Autonomous, algorithmic and reflective minds

Our cognitive architecture is based on on Stanovich's tripartite framework (Stanovich, 2010). Letters in this section refers to letters on Figure 1. Stanovich's tripartite model is a unified model of cognition that gives an account of how automatic (implicit) processes and explicit processes (control - attention and executive functions and more abstract planning and reasoning functions) are able to coexist. Stanovich's tripartite framework belongs to the "dual-process theories" we previously described System 1 (called "Autonomous Mind" in Stanovich's tripartite framework), is the locus of fast and automatic reasoning where instinctive behaviour, over-learned process, domain-specific knowledge, emotional regulation and implicit learning are found. System 2 is the locus of abstract and hypothetical reasoning. The division of human cognition into three sets of processes, instead of the traditional two of dual-process theories, provides a better account of individual cognitive differences. System 2 in Stanovich's tripartite framework is divided in two classes of processes, respectively called the "Algorithmic Mind," responsible for cognitive control, and the "Reflective Mind," responsible for deliberative processes. Owing to this subdivision, the framework can capture the distinction between on the one hand cognitive ability and fluid intelligence achieved by the algorithmic mind, and on the other hand thinking dispositions and critical thinking skills achieved by the Reflective mind.

The Algorithmic mind sustains three distinct processes: (1) override of Autonomous Mind processes (A), (2) cognitive decoupling (see section 2.1.2 for a complete description) (C), and (3) serial associative cognition.

Algorithmic Mind is linked to cognitive functions such us cognitive control and working memory and has access to information from the Autonomous Mind via two sets of pre-attentive processes (G) (1) perceptual processes and (2) beliefs and memory retrieval processes. Operations supported by the Reflective Mind define the subject's cognitive style. The Reflective Mind initiates: the override of Autonomous Mind (B) processes by the Algorithmic Mind and the cognitive decoupling operation (D) (the cognitive decoupling capacity is sustained by the Algorithmic Mind). The Reflective mind can also act upon the Algorithmic mind by interrupting (F) serial cognition to send a new action plan for execution or start cognitive decoupling.

2.1.2. Cognitive decoupling

Cognitive decoupling is a key mechanism that supports human rationality. Individual differences in the operation of this mechanism lead to differences in rational thinking (Stanovich 2010).

Decoupling has been largely studied in the dual process literature. It has been referred to as decoupling, cognitive simulation (Stanovich 2010) or hypothetical thinking (Evans 2008). It consists in the creation of temporary models (D) of the world upon which alternative scenarios can be experimented. Nichols and Stich (2000) dubs it "possible words box", a separate box in which simulation are carried out. The particularity of these temporary models is to be independent of the current mental representation of the world (primary representation). This prevents the real world representation to be confused with imaginary situations (secondary representation), since their manipulation doesn't affect the current representation of the world.

While the Algorithmic mind carries out cognitive decoupling, it has a hard time performing other processes. Decoupling is a cognitively expensive operation, (its cognitive load is higher than that of serial associative cognition) therefore it is not systematically performed, and when performed, it can be carried out incompletely. As a result, suboptimal responses that are cognitively easier solution provided by serial association (i.e. simple and incomplete models that appear appropriate for the situation) are often applied. Serial associative cognition (E) supports the implementation of these simple models.

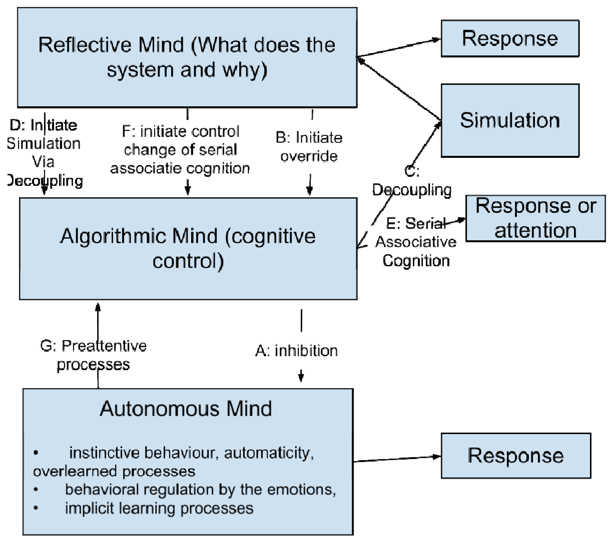

Figure: Stanovich's tripartite framework

2.2. Wisconsin Card Sorting Task

The Wisconsin card sorting test (WCST) (Grant & Berg, 1948) is a task widely used to test executive functioning, especially cognitive flexibility and abstract reasoning. The subject is shown a set of target cards. The figures on the cards vary in shape, number and color. Stimulus cards are shown to the subject, one by one, and the subject is asked to match them to one of the target cards. The subject is not told what the sorting rule is; he has to discover it. However, he receives negative feedback when his matching is wrong. After a number of successful matching, the rule changes without the subject being warned. After discovering the new rule, subjects experience difficulties inhibiting the previous rule: they cannot shift attention from one complex visual stimulus to another (set shifting). The perseverative error is the tendency to use the previous rule after a switch (Nyhus & Barcelo, 2009). Perseverative errors are more common in patients with lesions to the frontal lobe and increase with the age of subjects. The selection of the response is achieved through two mechanisms: the attentional set shift and the reversal shift, which are located in two distinct neuroanatomical structures of the PFC (Nagahama et al., 2005). The reversal shift, achieved through the posterior region of the PFC, is in charge of the update of associations between stimulus and response modules. The attentional set shift, achieved through the rostrodorsal PFC, is a higher level cognitive mechanism allowing adoption of the new rule by cognitive control (Anterior Cingular Cortex).

2.3. Computer simulations of the WCST

Computer simulations of this task have been achieved, mainly with connectionist approaches (Deahaene & Changeux, 1991; Kaplan et al., 2006), but also with symbolic and hybrid approaches (Kimberg & Farrah, 1993).

Dehaene and Changeux's simulation focuses on the functional aspect of the task by including in their neural network model the three cognitive components they deem critical to the accomplishment of the task: the ability to change rules when negative reward occurs, the capacity to memorize previously tested rules and the possibility of rejecting rules because of a priori reasoning. These three components are achieved by means of a hierarchical structure: a sensorimotor loop, a higher level assembly of rule-coding cluster codes – the current rule – shifted in case of negative reward (episodic memory of the system), an endogenous auto-evaluation loop allowing the internal testing of rule. The structure is compatible with the organization and specialization of cortical areas.

To simulate the distinction between the hypothesis generator and action, Kaplan et al. (2006) connected two distinct subsystems (respectively a hamming block and a Hopfield network), which allowed them to study perseverance and distractibility and reproduce lesions in the prefrontal cortex. The Hopfield network acts as the system's Working Memory, and the hamming block as its hypothesis generator.

Kimberg and Farrah (1993) explicitly represent the sorting behaviour of their system, to study the involvement of Working memory in the task, in an ACT-R architecture (Anderson & Lebiere, 1998). ACT-R is a production system and as such, uses a set of productions (its procedural knowledge and a working memory representation). The Working memory associations are weakened in order to model the effects of Frontal Lobe Damage in humans.

Being an hybrid approach, our architecture presents some similarities with these various architectures, if only because they all use a model of the human brain in the performance of the task. The endogenous auto-evaluation loop of Dehaene and Changeux's architecture, allowing the internal testing of rule, can be compared functionally to our decoupling ability. Their hierarchical structure can be compared to that one of our own architecture; however in our system, behaviour emerges through the interaction of the different levels, rather than being strictly controlled by a higher level structure. We share aspects with the different connectionist approaches since memory of the previous correct and incorrect rules is kept by the degree of activation of knowledge (at the reactive level) and objectives (at the reflective level). We share aspects with the symbolic approaches since knowledge and objectives in our system are represented by symbols. All of the previous simulations study the involvement of one specific cognitive component in the performance of the task. In this paper, however, we chose the WCST as a way of studying the interaction between the components (or minds) in Stanovich's Tripartite Framework, this task being known to tap reflective and algorithmic processes (inhibition and decoupling at the rule discovery stage).

3. OUR PROPOSAL

3.1. Architecture

We implemented our cognitive architecture in the multi-agent system (MAS) platform Madkit. The MAS is organized into groups (In Madkit, a "group" is a set of agents that share common characteristics), each corresponding to a cognitive level in our architecture. As there are three levels in our cognitive architecture (reactive, algorithmic and reflective), the MAS is composed of three groups. To keep matters simple, groups are given the name of the level they correspond to. Each agent has one or more roles (in Madkit, a "role" is an abstract representation of an agent's functionality) and belongs to one or more groups. All agents work in parallel, sending messages to each other. Each message received by an agent will activate the agent a little; the activation of the agent is thus a function of the number of messages it receives. The activation of an agent determines the number of messages it sends by unit of time.

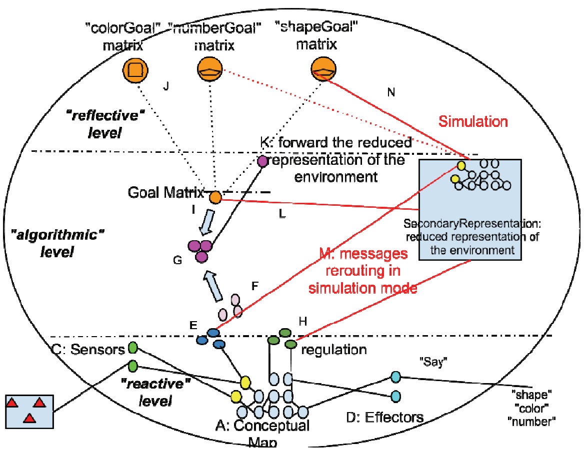

Figure 2 : System's architecture

3.1.1. Reactive level

Corresponding to Stanovich's Autonomous Mind, the Reactive level contains agents assigned with one of three roles: the "sensor" role (C - Letters in the section refer to letters in Figure 2), the "effector" role (D) and "knowledge" role (A), which we thus respectively name: Sensor Agents, Effector Agents and Knowledge Agents. Knowledge Agents form a network isomorphic to the conceptual map the system is initialized with. Each Knowledge Agent receives, as an additional role, a word from the conceptual map (e.g., "Red") and is linked to other Knowledge Agents according to links between the words in the map (e.g., the Knowledge Agent "Red' will be linked to the Knowledge Agent "Tomato", and so on). The links between Knowledge Agents are weighted according to the semantic distance between them in the conceptual map (e.g. the link between "Red" and "Tomato" will receive a higher weight than that between "Red" and "Meat"). The network of knowledge Agents makes up the system's declarative knowledge (semantic memory). In the experiments reported in this paper, the network of Knowledge Agents MAS was initialized with the common sense knowledge database Conceptnet (Havasi, Speer & Alonso, 2007). The number of messages exchanged between the agents, and therefore their activation, is at first determined by the distance between them in the conceptual map (it will also be determined by activation signals from higher levels – see below). Activation thus spreads through the network of Knowledge Agents (a process similar to semantic memories; Anderson, 1983).

Sensor Agents are sensitive to conditions in the environment (colors, words, numbers, shapes, etc.) and send messages to selected Knowledge Agents. For instance, a Sensor Agent that can detect colors (here: the recognizeColor agent) will be linked to color Knowledge Agents; it will send a message to the "Red" Knowledge Agent if it detects red in the environment. The system's environment is similar to (portions of) human environments. In the WCST simulation described below, the system is presented with cards identical to those human subjects see in real WCST experiments. Effectors Agents act on the environment. They are linked to Knowledge Agents whose role is relevant to the action they can perform. For instance, the NameColor Effector Agent will be linked to every color Knowledge Agent (thus implementing the capacity to name the colors we see). Taken together, the Sensor and Knowledge Agents make up the system's perceptual system. This means that the system's perceptual abilities are always a function of the Sensor Agents' information extracting capacities and of the system's knowledge about the environment, as implemented in the network of Knowledge Agents. Similarly, the Effector and Knowledge Agents form the system's action system. Both perception and action are thus fully situated and contextual. The system's long-term memory is made up of the network of Knowledge Agents in the Reactive Group, and the system's working memory (WM) at a given time is made up of the Knowledge Agents that are activated at that time. This implementation of working memory is consistent with the work of Engle (2010), in which WM is seen as a set of temporarily activated representations in long-term memory.

3.1.2. Algorithmic level

Corresponding to Stanovich's Algorithmic Mind, the Algorithmic Group is responsible for the control of the system. Based on a general idea developed by Cardon (2005), control is achieved by means of assessing and affecting the "morphology" of the system. RequestStatus Agents (E), which belong to both the Reactive and Algorithmic Groups, regularly query the status of Knowledge Agents; that is, number of messages sent to other agents during the last interval. Status Agents (F) represent the activity of the Reflective Group at a given time in the form of a *distance matrix* that records the message passing activity at that time (Status Agents will also send a reduced representation of the activity in the Reactive Group to the Reflective Group; see below). The distance between two concepts in the conceptual map is measured by the number of messages sent between the Knowledge Agents that bear the corresponding words as their role. Globally, this matrix thus represents a form or shape. The Algorithmic Level also contains the short-term goals of the system in the form of a graph of Goal Agents (which is sent by the Reflective level; see below). Each Goal Agent (I) contains a distance matrix that specifies the distance between each Knowledge Agents that is necessary if the system is to reach this goal.

Delta Agents (G) compute the difference between the matrix provided by the Status Agents and that provided by the Goal Agents. The resulting difference (another matrix) is provided to Control Agents (H) that send regulation messages to Agents in the Reactive Group, telling them to modify (i.e., increase or decrease) their activation so that their global activity more closely matches the shape of the current short-term goal. By activating elements of the system's long-term memory in relation to its current goal, thereby determining the current content of working memory, agents in the Algorithmic Group constitute the system's attentional system.

3.1.3. Reflective level

Corresponding to Stanovich's Reflective Mind, the Reflective Group is responsible for the logical and analytical skills of the system. Each agent in the Reflective Group has a shape (a distance matrix) as its role, which, as explained above, indicates the shape the Reflective Level must be in for the system to achieve a simple goal. Goal Agents (I) are organized in a directed graph. Every path in this graph represents a plan the system can applied to achieve a complex behaviour. Goal Agents are organized into a graph where each path represents a complex plan or strategy decomposed into a sequence of simple objectives (steps in the plan). A path (a sequence of simple goal path – in the WCST the path consist in a single objective but it could be longer for other tasks) (J) will be sent to Goal Agents of the Algorithmic Group, which will take care of its execution. Following Stanovich's Tripartite Framework, agents in the Reflective Group have access to a reduced representation of the environment, which, as explained, is provided as a matrix by the Status Agents of the Algorithmic Group (K). The similarity between these two matrices, computed by the Goal Agents, determines the activation of the Goal Agents, which propagates from the Agent most matching the reduced representation to those that follow in its path. The last agent in the path sends the parsed path to the Algorithmic Group. The shortest path (the simplest model) or the one the most activated (the model used more recently or more often) thus prevails over the other paths. Goal Agents of the Algorithmic Group will execute this path step by step (this corresponds to Stanovich's serial associative cognition).

The path executed by serial associative cognition provides the system with the sequentiality necessary to achieve complex goals. However, the system does not thereby lose its dynamicity. Reduced Representation of the environment are sent on a regular basis to the Reflective Group that can, based on the current state of the environment, interrupt serial cognitive association either by setting a new starting point in the path or by taking a new branch in the graph,. Decision-making at the Reflective level is therefore dynamically influenced by the current strategy (the decided path) of the system and the state of the environment.

3.1.4. Decoupling

Cognitive decoupling is an operation that is initiated by Agents of the Reflective Group and achieved by agents of the Algorithmic Group when multiple strategies (meaning two or more GoalSet Agents) are selected at the algorithmic level. The Goal Agent, which usually carries the unique goal matrix selected at the Reflective level triggers cognitive decoupling.

When the Goal agent hesitates between two strategies (when the activation levels of two GoalSet agents are close – closeness being defined by a sensitivity degree which we will further explain in section 4.4.1), it sends (L) a message to agents of the reactive group informing them that the system is now in simulation mode. It also triggers the creation of a possible world. A limited number of agents (currently 20) mirrors the activity of agents at the reactive level according to the reduced representation previously sent by Delta agents. Agents in this possible world are assigned dynamically the same roles and links as those agents from the Reactive Group they are replicating. This possible world corresponds to the separate secondary representation we presented earlier in section 2.1.2. Accordingly, SecondaryRepresentation agents are used instead of Knowledge agents and a distinct group (Algorithmic instead of Reactive) is used, to ensure that action on this secondary representation doesn't impact the current representation of the world (i.e., Knowledge Agents from the Reactive Group).

When agents of the algorithmic group are informed that they are in simulation mode, they reroute (M) their messages to the SecondaryRepresentation Agents of the Algorithmic Group instead of the Knowledge Agents of the Reactive Group. Once the simulation is completed, the activation of Goal Agents is regulated accordingly at the Reflective level (N), therefore dictating the future course of action.

Cognitive operations (goal inhibition and selection) are carried out by the Control Agents, Delta Agents, and agents from the Reflective Group during cognitive decoupling, making it considerably difficult to sustain other activities in the system at the reactive level (cognitive cost of decoupling).

For the WCST, the decoupling has been set to perform opposite style of thinking: a chosen categorization rule (the one with the highest activation or a random one if activations are equal) is applied in the possible world and is thereafter negated (negative feedback from the environment), thus a new categorization rule (the alternative) emerges. Goal agent therefore sends activation messages to the GoalSet agent bearing the first activated rule, and half less activation messages to the one bearing the alternative rule: if the first rule is negated in the "real" world, the alternative rule will therefore be the second one to be activated.

3.2. Neurological plausibility

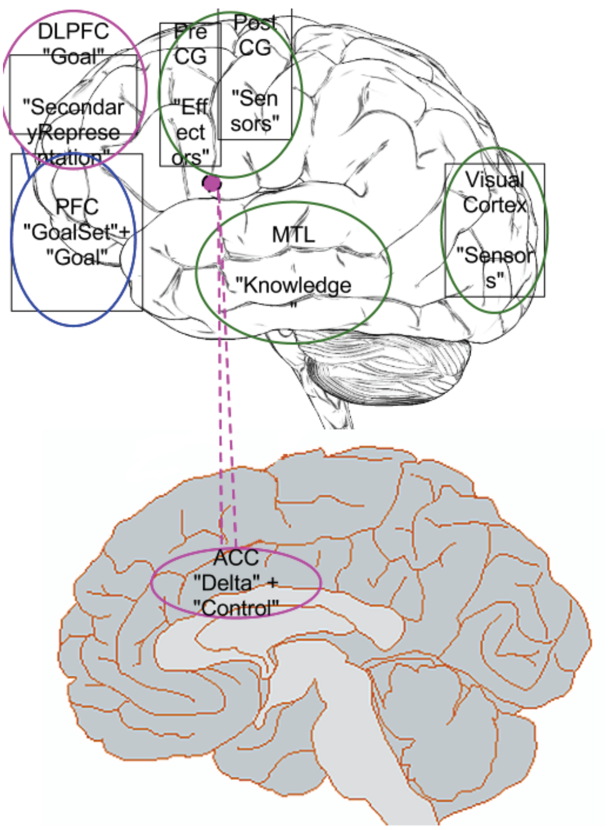

Figure 3: Functional mapping.

We cannot claim fine-grained neurological plausibility for this system; however, parallels can be drawn at the higher level of gross neurological structure and mesoscopic dynamical activity, allowing us to claim that a measure of neurological plausibility for the architecture. The neurological plausibility of dual process theories has been extensively studied (Goel, 2009; Lieberman, 2009). Stanovich's tripartite framework, individually, is also supported by neurological data (Stanovich, 2010). Since the design of our architecture is based on this model, it naturally inherits its neural plausibility. Cognitive tasks supported by the Algorithmic Mind lead to an activation of the Anterior Cingulate cortex (ACC). Performance of Algorithmic Mind processes leads to an activation of the ACC (Stanovich, 2010). There is furthermore evidence that decoupling (supported in our system by SecondaryRepresentation Agents) is achieved by the Dorsolateral PreFrontal Cortex (DLPFC) (Stanovich, 2010).

Furthermore, the different roles ascribed to the agents in the architecture correspond to functional roles that has been mapped to specific anatomical structures – see Figure 3 (Fuster, 2008; Botvinick et al., 2001). It must be noted that the ACC has been identified as the response conflicting monitoring system (Botvinick et al. 2001) in the human brain, regulating control's engagement. Conflict monitoring is achieved in our system by the collaboration between the Delta Agents and the Control Agents. The DLPFC, which provides support for goal directed behaviour is implemented in our system by Goal Agents. Regulation of posterior brain regions is implemented by the regulating messages sent by the Control Agents to agents of the Reactive Group. Knowledge Agents are linked to Sensor Agents as the Medial Temporal Lobe (MTL in figure 3) is known to mediate sensory memory. The medial temporal lobe is also identified as the functional locus of semantic memory. "Effector" and "Sensor" agents are associated with distinct roles in the Reactive Group since their functional role is achieved by distinct anatomical structures (PreCentral Gyrus, PostCentral Gyrus and Visual Cortex in figure 3). Through nested sensorimotor and goal-directing loops, we are therefore able to implement the cognitive dynamics of a goal-sensitive sensory-motor architecture.

Finally, we intend future developments of the architecture to further increase its neurological plausibility. The majority of the existing circuits in the system (its message passing channels) mirrors neurochemical pathways of the human brain (for example, dopamine pathway will in our system be represented by the message passing activity between GoalSet Agents and Goal Agents in the Reflective Group, and between Knowledge Agents and Effector Agents, to respectively represent the reward effect of dopamine, and its impact on motor function (Fuster, 2008)). This property will allow us to further explore the dynamics of cognitive processing in the system in future works by varying the message passing activity in distinct pathways (neuromodulation).

4. EXPERIMENT AND RESULTS

4.1. Implementation

We chose a “generic” multi-agent Platform (Madkit) that allows the creation of agents with different ranges of complexity and in which large numbers of agents can operate in parallel. Madkit implements the AGR (Agent/group/role) model, which we found particularly suitable to implement the various groups of agents (levels), and the diversity of agents in each group, as described previously. In AGR (Ferber et al., 2003):

- An agent is an active entity communicating and playing a role within one or several groups. There are no design constraints regarding the complexity of the agents (reactive/cognitive).
- A group is a set of agents sharing common characteristics. Groups define the or ganizational structure of the system.
- A role is an abstract representation of agent’s functionality within a group. We chose to assign a thread to each agent in order to preserve its dynamicity.

4.2. Experimental setting

The system was given as parameters three goals matrices corresponding to the color, shape and number categorization. To implement the context effect of the four reference cards that the subject must select to place each of his response cards (material adapted from Heaton et al., 1993), we added the following links in the conceptual map of the system: Red – triangle – one, Green – star – two, Yellow – square – three, Blue – circle – four. We also linked the shape, color, and number knowledge (already present in ConceptNet) to the sensors. A script provided the system with the series of cards it had to categorize and evaluated the system’s answer. The categorization rule was changed after 6 consecutive successes. The experiment consisted of 128 cards with figures varying in shape, number and color. The script attempted to test 6 categorization rules (“shape”, “color”, “number” x 2) on the 128 cards. No warning was sent to the system before a rule change.

4.3. Description of the interactions

Although our system is not sequential in nature (organizations work in parallel and the global behaviour of the system emerges from the interactions of all agents in the system), we describe below the WCST task processing sequentially to ease its understanding. However, it should be borne in mind, that this sequentiality emerges from the system’s parallel processing.

Table 1: Variation of decoupling’s degree of sensitivity.

Degree of sensitivity	1/6	1/8	1/10
Trials	89.1	120.41	126.6

Table 2: Excerpt of one simulation’s logs

Serie	Trial	Response	Simulation
Color	1	correct	
	2	correct	
	3	correct	color shape
	4	Incorrect : shape	
	5	correct	

Step 1 - When a new card appears in the environment, sensors (for color, shape, and number) extract its relevant properties and forward the information to Knowledge Agents in the Reactive Group where activation spread according the links in the conceptual map.
Step 2 - Status Agents forward this information (activation) to the algorithmic level.
Step 3 - A reduced representation of the environment is produced by Delta Agents and forwarded to the reflective level
Step 4 – The received reduced representation leads to the activation of competing rules (by pattern matching between the goal they are bearing and information from the reduced representation).

(Please note that Step 5 to Step 7 – included – are conditional to the fact that there is more than one winning GoalSet Agent, otherwise, step 8 is directly applied.)

Step 5 - When there is more than one winning GoalSet Agent, meaning more than one winning goal, cognitive decoupling occurs. A mini world (secondary representation made after the – primary - representation of the world carried by the system at the reactive level) is created where agents are initialized after the reduced representation of the world sent at step 4.
Step 6 – Messages from the Control agents (regulation) and from RequestStatus agents (status of the concept agents) are rerouted from the reactive level to the secondary representation that has been initialized at step 5.
Step 7: Rules that emerged in the secondary representation during cognitive decoupling are sent to the reflective level (with different activations).
Step 8: The corresponding matrix (the one that a GoalSet Agent was carrying) is sent to the algorithmic level.
Step 9 – Rerouting is stopped. Agents of the Algorithmic group are back in charge of the regulation the Reactive level. They regulate agent’s activity

according to the activity matrix associated with the system's current goal.

4.4. Results

4.4.1. Cognitive decoupling

In the system, decoupling occurs when two competing answers are identified (when two answers have close activation levels). We ascribed a degree of sensitivity to the Goal agent which starts decoupling operations: the degree of sensitivity determines how close the two competing answers need to be to start cognitive decoupling.

In Table 1, we show the performance of the system in regards to this degree of sensitivity. We can see that the higher the degree of sensitivity is (meaning the more decoupling operation occur), the lower is the number of trials required to perform the task. Please note that, due to the experimental procedure (section 4.2), the maximum number of trials allowed to complete the task can't be higher than 128. For the "1/6" degree, system's performance was better than for human subjects, this results shows interesting premises for the usability of this architecture for the simulation of cognitive tasks with different cognitive profiles (replicating individual differences of subjects with different cognitive styles).

Table 2 presents the log for the simulation of a WCST task with the 1/8 degree. For the "color" serie, we can observe the decoupling being launched in the following situations:

- Trial 3: a simulation/decoupling is launched due to two competing answers ("color" and "shape"). The system creates via decoupling a possible world (separate box) where the color categorization rule is applied and ruled wrong. The second emerging rule is the shape categorization rule.
- Trial 1 and 2: "color" (a wrong answer) is selected. In the possible world created, "shape" activated first. The possible world being a copy of the system's representation of the world, "color" has already been marked as a wrong answer, therefore the selected second answer is number.

4.4.2. Factors analysis

We focused our analysis on 4 factors:

- Failure to maintain set: failure to carry out a complete categorization after a number of consecutive correct trials;
- Perseverative errors: Mean number of trials in which the subject persists in categorizing items with the preceding rule after rule change;
- Categories completed: mean number of completed categories;
- Trials: Mean number of trials to complete 6 categories.

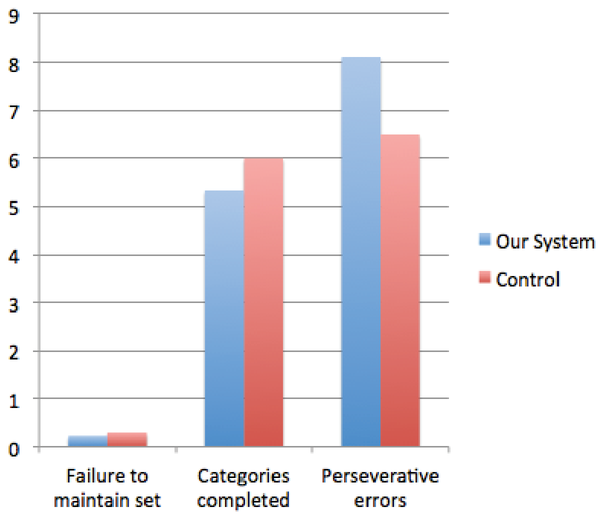

Figure 3: Mean « Failure to maintain set » errors, Perseverative errors and categories completed by our system (mean number for a hundred trials) and by a control group of healthy human subjects taken from Bishara *et al.*'s study (2010).

We compared our results (degree of sensitivity: 1/8) for the three first factors to those of human subjects. Results are presented in figure 3.

The fourth factor, mean number of trials to complete the task in our system was 120.41 against 101.12 in unpracticed healthy human subjects (Basso et al., 1999). The maximum number of rules the system (for his 128 trials) was able to discover and apply varied between 5 and 6 (mean number being 5.33).

During simulations, we were able to observe two different types of wrong answers: those due to a failure at the stage of rule-discovery (reflective level – failure at the hypothesis testing level) and those due to a failure of inhibition (inhibition of the reactive level by the algorithmic level – failure of the working memory in the human subject).

Our system was able to reproduce the pattern of results observed in human subjects (see Figure 3); however, its perseverative tendency was worse than that of human subjects. We believe that this result reflects a failure of inhibition on the part of Control Agents, showing a deficit at the algorithmic level: a failure to inhibit the routine activity that had been established by previous trials at the reactive level. On the other hand, the low value for the « failure to maintain set » factor shows a good flexibility of the system for the selection of new rules, therefore validating the work of the reflective level. To improve the system in future works, we therefore plan to work on calibrating the message passing activity of Control Agents in the Algorithmic Group. This correction might also help increase the number of completed categories.

5. CONCLUSION AND FUTURE WORKS

In this paper, we present a cognitive architecture implementing Stanovich' Tripartite Framework. We already produced classical and semantic Stroop task simulations to validate the two first levels of the cognitive architecture: algorithmic and reflective mind (Larue et al. 2012). In this paper, we present a validation of its computational soundness as a whole. The WCST simulation we perform on this architecture is a concrete demonstration of how the three levels can interact to produce a complex behaviour involving different levels of cognition in one single structurally unified tool. The WCST allows a study of the interaction between the algorithmic and reflective levels, and more specifically the initiation of simulation/decoupling processes. Decoupling is a key process that sustains deliberative behaviour by allowing hypothesis testing. In this task, automatic cognition is observed in the initial reaction of the system to stimuli, algorithmic cognition allows the system to achieve trial-and-error adaptation and hypothesis testing (e.g. simulation) initialized thanks to the Reflective level from which will emerge the action to be performed by the system. To implement the dual nature (sequentiality/rule-following and reactivity/dynamicity) of human cognition, and thus meet the duality challenge, we combine aspects of classical symbolic (sequentiality/rule-following) and connectionism (reactivity/dynamicity) into a structurally unified cognitive architecture. We were able to combine those two approaches (dynamicity and sequentiality) using one single computational intelligence paradigm (Multi-agent system). Results were encouraging concerning the interaction of reflective and reactive levels of the system, but they we will need in the future to improve the system's inhibition ability at the algorithmic level..

In future work, we plan on addressing this issue by introducing a new variable to the system: neuromodulations. Neuromodulations would modulate communications between agents inside a group and between agents in different groups of the system, therefore modifying the system's general dynamics (allowing us finer-grained control on the dynamics but also, for example, to induce pathological behaviour in the system). Also, this task enabled us to test hypothesis testing abilities which are primordial for a deliberative behaviour (especially selection of the good strategy); however, the strategies among which the system had to chose where very simple strategies. In the future work, we plan on using the system with more complex strategies.

REFERENCES

Anderson, J. R. (1983). The architecture of cognition. Cambridge, Mass.: Harvard University Press.

Basso, M.R., Bornstein, R.A., & Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. The Clinical Neuropsychologist, 13.

Bishara, AJ, Kruschke, JK, Stout, JC, Bechara, A, McCabe, DP, Busemeyer, JR (2010), Sequential Learning Models for the Wisconsin Card Sort Task, J Math Psychol.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., et Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev, 108*(3), 624-652.

Carruthers, P. (2006). The Architecture of Mind. Oxford: Oxford University Press.

Cardon, A. (2005). La complexité organisée. Paris: Hermès. Dehaene, S., & Changeux, J. (1991). WCST: Theoretical

analysis and modeling. Cerebral Cortex, 1

Dehaene S. et Changeux, J., «The Wisconsin Card Sorting Test: Theoretical analysis and modelling in a neuronal network », Cerebral Cortex, 1991.

Eliasmith, C. (2005). Cognition with neurons: A large-scale, biologically realistic model of the Wason task. In L. B. Bara G.; Barsalou, M. (Ed.), *Cognitive Science 2005.*

Engle, R. W. (2010). Role of Working Memory capacity in Cognitive Control. Current Anthropology, 51.

Evans, J. (1984), Heuristic and analytic processes in reasoning. British J of Psych, 75.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. Psychonomic Bulletin and Review. 13(3), 378-395.

Evans, J. (2008). Dual-processing accounts of reasoning, judgement and social cognition. An Rev Psy, 59.

Ferber, J., Gutknecht, O., et Michel, F. (2003). *From Agents to Organizations : an Organizational View of MultiAgent Systems*. Paper presented at the Agent-Oriented Software Engineering (AOSE) IV, Melbourne.

Fodor, J. (2000). The Mind Doesn't Work That Way:The Scope and Limits of Computational Psychology. MIT Press.

Frankish, K., et Evans, J. S. (2009). The duality of mind. In Frankish & Evans (Eds.), In Two Minds.

Fuster, Joaquin M. (2008). The Prefrontal Cortex, Fourth Edition. Boston: Academic Press.

Goel, V. (2009). Cognitive Neuroscience of Thinking. In G. Berntson & J. Cacioppo (Eds.), Handbook of Neuroscience for the Behavioural Sciences: Wiley

Grant, D. A., et Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting in a Weigl-type card-sorting problem. J of Exp Psych, 38.

Havasi, C., Speer, R., et Alonso, J. (2007). ConceptNet 3. Natural Languages Processing 2007.

Heaton, R., Chelune, G., Talley J., Kay, G., Curtiss (1993) Wisconsin Card Sorting Test Manual: Revised and expanded. Psychological Assessment Resources, Odessa, FL.

Kahneman, D. (2011) Thinking, fast and slow. Farrar, Straus and Giroux.

Kaplan, G.B., Sengör, N.S., et al. (2006). A composite neural network model for perseveration and distractibility in the WCST. Neural Networks, 19.

Kimberg, D.Y. and Farah, M.J. (1993). An unified account of cognitive impairments following frontal lobe damage. Journal of Exp. Psych., 122.

Larue, O., Poirier, P., Nkambou, R. (2012). A Three level Cognitive architecture for the simulation of cognitive phenomena. CAI (Canadian Artificial Intelligence)

Lieberman, M. (2009). What zombies can't do : a social cognitive neuroscience approach to the irreducibility of reflective consciousness. In K. Frankish & J. S. Evans (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 293-316): Oxford University Press.

Nagahama, Y., Okina, T., Suzuki, N., Nabatame, H., et Matsuda, M. (2005). The cerebral correlates of different types of perseveration in the WCST.J Neuro Psych, 76(2).

Nichols, S. and Stich, S. (2000) "A Cognitive Theory of Pretense." Cognition, 74, 115-147.

Nyhus,E., Barcelo,F.(2009).WCST and the cognitive assessment of prefrontal executive functions. Brain Cogn, 71(3).

Sherry, D. F. and Schacter, D. L. (1987). The evolution of multiple memory systems. Psych Rev, Vol 94(4).

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22

Sloman, A. Chrisley, R.(2005). More things than are dreamt of in your biology: Information-processing in biologically inspired robots. Cognitive Systems Research 6(2).

Stanovich, K. E. (2010). Rationality and the reflective mind: Oxford University Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Intelligence and rationality. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge Handbook of Intelligence* (pp. 784-826). New York: Cambridge University Press.

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

Sun, R. (2004). Desiderata for cognitive architectures. Philosophical Psychology, 17(3).

AUTHORS BIOGRAPHY

Othalia Larue is a PhD student in Cognitive Computer Science at the Université du Québec à Montréal, and member of the GDAC (Knowledge Management Research) Research Laboratory. Her research interests include cognitive simulations, cognitive architectures and affective computing.

Pierre Poirier (Ph.D.) is a professor of philosophy at the Université du Québec à Montréal and director of its Cognitive Science Institute. His interests are the philosophy of science (mainly of cognitive science and of neuroscience).

Roger Nkambou (Ph.D.) is a professor of Computer Science at the University of Quebec at Montreal, Director of its Cognitive Computer Science doctoral program and Director of the GDAC (Knowledge Management Research) Laboratory (http://gdac.dinfo.uqam.ca). His research interests include knowledge repre- sentation, intelligent tutoring systems, intelli- gent software agents, ontology engineering, student modeling and affective computing.