Biologically Inspired Cognitive Architectures (2013) 6, 67–75

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/bica

RESEARCH ARTICLE

Hypothetical-thinking based on cognitive decoupling and thinking dispositions in a dual cognitive agent

Othalia Larue *, **Pierre Poirier**, **Roger Nkambou**

GDAC Research Laboratory, Université du Québec à Montréal, Canada

Received 16 April 2013; received in revised form 24 June 2013; accepted 28 June 2013

KEYWORDS
Cognitive decoupling;
Dual process theories;
Wason card selection task;
Hypothetical-thinking

Abstract
We present an implementation of the hypothetical-thinking capacities of our cognitive architecture, based on Stanovich's tripartite framework (Stanovich, 2009). To illustrate and study processing by this new feature, we simulated a well-known task in the psychology of reasoning (the Wason card selection task) with four different cognitive styles (strongly reactive, purely executive, weakly reflective, purely reflective) and were able to reproduce the results and types of errors found in studies of human reasoning abilities. The first three profiles account for the results of 90% of human subjects (all those who provide answers that are not acceptable by logical standards). The strongly reactive profile gave a plausible account of the way humans provide the logically incorrect answer to the task. The purely executive and weakly reflective gave an account of how subjects can provide part of the correct answer. The last profile (purely reflective), a much a much slower process, produces a complete and correct answer by logical standards. While the purely reflective process was the only one able to do this, it is to be noted that the purely executive system was able to provide a correct (but incomplete) answer using less computational resources (time).

Introduction

Evidence from many fields in cognitive psychology (see Evans (2008) for a review) and cognitive neuroscience (see e.g. Goel, 2009) supports the view that human minds are composed of two systems with functionally incompatible features: (1) dynamical and reactive, and (2) sequential and rule following. To build a unified cognitive architecture that reproduces these, we used a dual-process theory for our architecture: Stanovich's tripartite framework (2009) which gives an account of how reactive and reflective behaviours emerge through the interaction of three ''minds'': the autonomous, algorithmic and reflective minds.

Corresponding author. Tel.: +1 335149873000.
E-mail addresses: othalia.larue@gmail.com, larue.othalia@courrier.uqam.ca (O. Larue).

2212-683X/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bica.2013.06.004

While we focused in earlier work (Larue, Poirier, & Nkambou 2013) on the interactions between automatic and attentive processes, we focus here on the higher level processes that initiate and sustain hypothetical thinking through the building and manipulation of off-line simulations of the real world. We illustrate the system's processing with a standard task in the psychology of reasoning, the Wason card selection task (Wason, 1966), which we chose because it typically illustrates the strengths and weaknesses of human reasoning abilities. However, rather than looking at the reasoning abilities in a evaluative fashion (at strengths and weaknesses), we see them in this paper as different thinking styles, which lead to partial or complete solution given the cognitive resource (and hence cost) engaged by the individual, or here the system. A cognitive architecture that displays the same reasoning profile as humans is necessary, we believe, to meet the BICA challenge.

Related work

The Wason card selection task

The Wason task (see Fig. 1) is one of the most popular tests in the psychology of reasoning. Four cards are presented to the subject, who is informed that, for each card, there is a letter on one side and a number on the other.

Since the rule to be verified is a conditional statement, (classical) logic states that subjects should choose to turn the card with a ''A'' and the card with a ''7''. However, the performance is quite low on this task: only about 10% of subjects provide the right answer. While most subjects correctly offer to turn the ''A'' card, most also do one of two mistakes (sometimes both). Some fail to offer to turn ''7'' card, even though if there was an ''A'' is behind the that ''7'', then the rule to be verified would be invalidated. Many also have the tendency to select the ''3'' card instead, although the presence of something else than A on the other side of this card would not invalidate the rule. In this task, there are thus two types of logical errors: (1) failing to select the required ''7'' card and (2) selecting the non-required ''3'' card. The first error shows that subjects, while they apply the *Modus Ponens* rule (the truth of the antecedent implies the truth of the consequent), they fail to apply the equivalent *Modus Tollens* (the falseness of the consequent implies the falseness of the antecedent). The second error shows that subjects are overly influenced by the wording of the statement.

A dual-process account of the Wason task results

The heuristic-analytic theory (Evans, 1989) gives an account of the various results from the Wason task. It explains the errors encountered in that task (as well as in other psychological tasks) by positing the presence of a ''heuristic'' (fast reasoning) and an opposing ''analytical'' (slow reasoning) systems. Only this latter system can, it is posited, deal with abstract and hypothetical reasoning.

Evans' theory is based on three principles (Evans, 1998): the singularity principle, the relevance principle, and the satisficing principle. The singularity principle states that, due to our cognitive system's limited capacity, only mental models that represent a single hypothetical situation are constructed. The relevance (or heuristic) principle states this mental model is constructed pragmatically with respect to the context. The Information needed to construct the model is selected from relevant information found in memory, where relevance is judged by the information's credibility and probability. The constructed mental model is then evaluated by the analytical system according to a satisficing principle according to which subjects will entertain a mental model that may be biased as long as they do not find a reason to abandon it. Biases may occur because the heuristic system does not see the subject's logical knowledge (his knowledge of logical rules for instance) as relevant to solve the task. The subjects' performance is victim of a matching bias: linguistic elements in the wording of the task trigger the use of heuristics. The presence of the conditional (''if'') triggers the use of the if-heuristic, which selectively directs the subject's attention on the main items in the conditional statement's antecedent and consequent, thus creating a perceptual matching bias where the subject matches his answer to the task's wording. By contrast, in social relations contexts or when the statement of the task has deontic content, subjects' performance is greatly increased. This is probably because relevance is no longer determined by linguistic factors but rather by pragmatic factors, which leads to the disappearance of the matching bias (Evans, 2006). We focus in this paper on the abstract version of this task, as we find it more appropriate to study our architecture's reflective processes.

Stanovich's tripartite framework (Stanovich, 2009) is a specific dual-process theory we used as a unified theory of cognition to build our cognitive architecture. It provides an explanation of how reflective (characterized by sequentiality) and adaptive (characterized by reactivity) human behaviour emerges from the interaction of three distinct cognitive levels (which he calls minds): (1) an autonomous/reactive mind responsible for fast context-sensitive behaviours, (2) an algorithmic mind responsible for cognitive control, and (3) a reflective mind responsible for deliberative processing and rational behaviour.

Although in general agreement with Evans' heuristic-analytic account of the Wason task results, Stanovich emphasizes that, even when subjects fail at the task, some

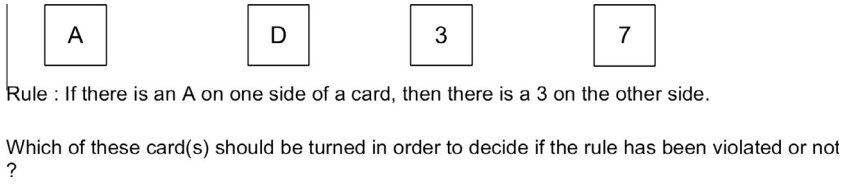

Fig. 1 The Wason task.