Journal of Applied Research in Memory and Cognition 4 (2015) 180–190

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Original Article

A general instance-based learning framework for studying intuitive decision-making in a cognitive architecture

Robert Thomson*, Christian Lebiere, John R. Anderson, James Staszewski

Carnegie Mellon University, United States

ARTICLE INFO

Article history:
Received 22 October 2013
Accepted 11 June 2014
Available online 19 June 2014

Keywords:
Cognitive architecture
Cognitive modeling
Decision-making
Heuristics

ABSTRACT

Cognitive architectures (e.g., ACT-R) have not traditionally been used to understand intuitive decision-making; instead, models tend to be designed with the intuitions of their modelers already hardcoded in the decision process. This is due in part to a fuzzy boundary between automatic and deliberative processes within the architecture. We argue that instance-based learning satisfies the conditions for intuitive decision-making described in Kahneman and Klein (2009), separates automatic from deliberative processes, and provides a general mechanism for the study of intuitive decision-making. To better understand the role of the environment in decision-making, we describe biases as arising from three sources: the mechanisms and limitations of the human cognitive architecture, the information structure in the task environment, and the use of heuristics and strategies to adapt performance to the dual constraints of cognition and environment. A unified decision-making model performing multiple complex reasoning tasks is described according to this framework.

This article describes how computational models of intuitive decision-making are expressed within the constraints of the *ACT-R cognitive architecture* (Anderson et al., 2004). These models are noteworthy for their ability to explain a variety of heuristics and biases in terms of the processes and representations that produce them. These phenomena have largely been captured and defined as results of experimental manipulations (Kahneman & Tversky, 1996) but not in terms of process models justified by a cognitive architecture (Dimov, Marewski, & Schooler, 2013). A concern of modeling intuitive decision-making behavior using cognitive architectures is confounded by the explicit decisions encoded by the modelers. This criticism can be described as: instead of modeling intuitive behavior per se, cognitive models make explicit the intuitions of their designers (Cooper, 2007; Lewandowsky, 1993; Schultheis, 2009; Cooper, 2007; Lewandowsky, 1993; Schultheis, 2009; Shallice & Cooper, 2011). We address this criticism by showing that the instance-based learning mechanisms in the ACT-R cognitive architecture (Gonzalez, Lerch, & Lebiere, 2003) exhibit the characteristics of intuitive decision-making as described in Kahneman and Klein, (2009), and provide a clearer distinction between automatic and implicit (System 1) processes and deliberative and explicit (System 2) processes. In addition, we specifically address this *modeler selection* criticism by showing that the explicit strategies of the models instantiate the theories of the model designer and thus are a mechanism for theory evaluation rather than a confounding factor in model development.

In making this argument, we recommend adopting a tripartite explanation of decision-making and biases that illustrates the critical role of the task environment in the decision-making process. We argue that decision-making should be understood in terms of: (1) the mechanisms and limitations of the architecture; (2) the information structure in the task environment; and (3) the use of heuristics and strategies to adapt performance to the dual constraints of cognition and environment. From examples of existing models, we show that simulating behavior within a cognitive architecture is a useful methodology for the study of the mechanisms, variables, and time-course in complex decision-making processes that are impossible in experimentation due to exploding combinatorics.

1. What is intuitive decision-making?

Simon (1992) characterized intuitive decision-making skill as "nothing more and nothing less than recognition" (p. 155). In their seminal work on expertise, Chase and Simon (1973) identified that chess experts require upwards of a decade of study

* Corresponding author. Tel.: +1 4126890536.
E-mail addresses: thomsonr@andrew.cmu.edu, noblephoenix@gmail.com (R. Thomson).

http://dx.doi.org/10.1016/j.jarmac.2014.06.002

to retain 50,000–100,000 distinct and rapidly accessible patterns of chess positions. Intuitive decision-making has been studied in both the *naturalistic decision-making* and *heuristics and biases* literature, with the former generally focused on the successes of intuitive reasoning, while the latter generally focused on its failures (Kahneman & Klein, 2009). A distinguishing feature of intuitive decision-making is that a single plausible solution rapidly 'comes to mind' in its entirety without explicit or conscious awareness of the causal factors entering into the decision (i.e., not being consciously derived in a piecemeal, step-by-step, or in a 'deliberative' manner; Newell & Simon, 1972; Simon, 1995). As such, intuitive reasoning is considered System 1. For example, Klein, Calderwood, and Clinton-Cirocco (1986) found that fire marshals tended to make rapid decisions by generating a single alternative, mentally simulating its outcome, and either making minor revisions or adopting the next closest alternative. Effectively, fire marshals were pattern-matching based on their prior experiences. This strategy has been termed *recognition-primed decision-making*.

Conversely, deliberative decision-making is often characterized as strategic, effortful, slow, and rule-oriented (Klein, 1998), and as such is considered System 2 thinking (Kahneman & Frederick, 2005; Stanovich & West, 1999). Interestingly, the act of verifying an intuition is generally seen as optional, effortful, and thus a function of System 2 (Kahneman & Klein, 2009).

In order to gain intuitive expertise, two conditions first need be met. The first condition is that people receive extensive practice in a task environment that is sufficiently stable and provides causal or statistical cues/structures that may at least theoretically be operationalized (Hogarth, 2001; Hogarth, 2001; Brunswik, 1957). This need not be deterministic (e.g., playing poker is a probabilistic but stable environment; Kahneman & Klein, 2009). The second condition is that there must be sufficient feedback from the task environment which provides people an opportunity to learn the relevant cues/structures. In other words, feedback must be sufficient to generate a relevant internal problem space. This requirement of feedback and interaction with the task environment drove our adoption of the tripartite level of description.

2. Why use a cognitive architecture?

Cognitive architectures model behavior using a set of common mechanisms and processes (i.e., *the architecture*) whose goal is to not only explain human behavior, but the underlying structures and representations subsuming cognition as a whole. These mechanisms should be both psychologically and neurally plausible to account for human behavior. This level of description is not generally captured by either mathematical or informal models of decision-making. Before getting into further details of mechanisms, models, and results; there is an important argument to be made for the role of cognitive architectures in general, which is best characterized by Herbert Simon (in 1971, no less):

> The programmability of the theories is the guarantor of their operationality, an iron-clad insurance against admitting magical entities into the head. A computer program containing magical instructions does not run, but it is asserted of these information-processing theories of thinking that they can be programmed and will run. They may be empirically correct theories about the nature of human thought processes or empirically invalid theories; [but] they are not magical theories. (p. 148)

In modern terms, simulations using a cognitive architecture provide a falsifiable methodology for the study of cognitive processes and representations, a particularly important characteristic when studying largely implicit processes such as intuitive decision-making. They serve several theoretical functions including: organizing and relating a substantial number of cognitive mechanisms, making testable predictions, and explaining the cognitive processes underlying human performance. In many cases, cognitive models can perform tasks too complex to analyze with traditional experimentation due to the combinatorics of the possible decision space. As will be described, a single ACT-R model has explained anchoring and adjustment, confirmation, and probability matching biases across a range of complex geospatial intelligence tasks using a common instance-based learning approach (Lebiere et al., 2013). Similarly, Marewski and Mehlhorn (2011) were able to specify 39 process models studied in decision-making using a smaller subset of 5 ACT-R models. In short, cognitive architectures allow for theories to be constrained by scientifically established mechanisms and (hopefully) easily describable processes.

This is not to argue that cognitive architectures are a panacea for studying decision-making (or psychology in general), but we do claim that they are a valuable tool in the generation and exploration of theories (c.f., models) which may be too complex for traditional piecemeal experimental methods. In particular, intuitive decision-making tends to be cognitively 'opaque' with little observable evidence, and what little evidence there is coming from highly fallible introspection. As such, many descriptions of intuitive decision-making are inherently qualitative or are characterized using relatively simple experimental results (Dimov et al., 2013). An advantage of cognitive architectures is not only their ability to objectively explain accuracy and response times in terms of the operation of both symbolic elements and their sub-symbolic activation strengths (and in the case of ACT-R, links to neural structure), but also the ability to go 'under the hood' and actually *look inside the model to explicitly examine causal processes*. Such computational cognitive models make testable predictions of what is going on *inside the mind* of someone performing intuitive decision-making.

One measure for validating *inside the mind* predictions is to perform model tracing. Model tracing is a technique where a model is forced to respond with some or all of the same values as a human participant, and then the internal states of the model are examined to determine the influence of these 'forced' decisions. By examining the commonalities between the model's internal states and human behavior, modelers are potentially able to make causal claims about the nature of mental processes within participants; that is, to explain how human performance is produced by various cognitive mechanisms and their interaction. This performance includes traditional measures such as accuracy and response time, but also predictions of fMRI bold response for specific brain areas associated with the functional modules of the cognitive architecture (Anderson, 2007).

The benefits of cognitive architectures can be seen as bridging or synthesizing formal mathematical theories (such as Bayesian modeling) and knowledge-level strategies (e.g., heuristics). As such, cognitive architectures act as a link between Marr's (1982) computational and algorithmic levels, with the benefits of a corresponding bridge to the physical level (i.e., neural) implementation. Bayesian models belong to a broad class of abstract models that formally (i.e., mathematically) explain human behavior in terms of processes computing probabilities over a set of possible decisions. While Bayesian (and related probabilistic) models do provide an explanation of behavior, it is not generally accepted to be a cognitively (i.e., psychologically) plausible one as the underlying mechanisms driving the processes are somewhat vague or not tractable (Bowers & Davis, 2012). As such, Bayesian theories belong at the computational level of Marr's hierarchy. This is not a criticism specific to Bayesian models, but can also be applied to other mathematical theories such as prospect theory (Kahneman & Tversky, 1979), decision theory (Berger, 1985), and quantum probability theory (Busemeyer, Pothos, Franco, & Trueblood, 2011). Similarly, explanations in the