Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 29–30 (2014) 1–30

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A model of personality should be a cognitive architecture itself

Action Editor: Vasant Honavar

Ron Sun *, Nicholas Wilson

Department of Cognitive Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Received 26 September 2013; received in revised form 16 December 2013; accepted 29 January 2014
Available online 19 February 2014

Abstract

This paper describes how personality may be explained by a generic, comprehensive computational cognitive architecture. We show that a cognitive architecture by itself can serve as a generic model of personality, without any significant addition or modification. A cognitive architecture can capture the fundamental invariance within an individual in terms of behavioral inclinations as well as the inevitable variability of behaviors. Various tests and simulations have been conducted within the cognitive architecture that show that such a model is reasonably stable, is relatively flexible (in terms of person–situation interactions), captures some major personality traits (e.g., the Five-Factor Model), and accounts for a variety of empirical data. The work shows the feasibility and usefulness of integrating personality modeling with generic computational cognitive modeling (i.e., cognitive architectures).

Keywords: Personality; Cognitive architecture; Cognitive modeling

1. Introduction

Personality, as extensively studied in personality psychology, should be the result of various existing psychological mechanisms and processes that have been commonly identified, and nothing else. That is, there should not be any special entities, mechanisms, or processes of personality (as has been argued before). Cervone (2004), for example, has argued that personality is a complex system with dynamic interactions among multiple processes; thus, personality should be understood by reference to basic cognitive processes that give rise to overt behavior. Shoda and Mischel (1998), Mayer (2005), and Sun and Wilson (2011) made similar points.

A computational "cognitive architecture" should, ideally, include all essential psychological entities, mechanisms, and processes of the human mind (Sun, 2004). The notion of a cognitive architecture should ideally be close to that of "personhood" (Pollock, 2008; Taylor, 1985). Within the cognitive architecture, the interaction among different subsystems (components and their mechanisms and processes) should be able to generate psychological phenomena of all kinds, which of course include personality-related phenomena (Sun & Wilson, 2011).

Thus, personality, if it is a valid psychological construct, should be accounted for by a cognitive architecture, without any significant additions or modifications of mechanisms or processes within the cognitive architecture. A cognitive architecture should have essential components, mechanisms, and processes of the mind (such as various memory modules, inference mechanisms, learning mechanisms, and so on) built-in. Otherwise, it would amount to a software tool, which allows one to build whatever models that one wants to build but does not sufficiently specify the architecture of the mind. In addition, it is desirable that a model of personality captures more detailed aspects of psychological processes than previous work, and goes beyond

* Corresponding author.
E-mail address: dr.ron.sun@gmail.com (R. Sun).

http://dx.doi.org/10.1016/j.cogsys.2014.02.001
1389-0417/

specialized computational models addressing personality alone (e.g., Shoda & Mischel, 1998). It is necessary to go beyond abstract notions of goals, plans, cognitive affective units, and so on—It is one thing to argue that personality traits consist of configurations of goals, plans, or cognitive affective units, it is quite another to map personality traits to more concrete and better grounded psychological processes and mechanisms. It is useful to ground personality traits in a cognitive architecture, so that they can be explained in a deeper and more unified way, along with many other psychological phenomena, based on the same primitives as envisaged within a generic cognitive architecture (without any significant additions or modifications). Also, coupled with the account of learning in the cognitive architecture, a model of personality should be able to account for the emergence, shaping, and tuning of personality by a variety of factors.

Therefore, personality should be computationally captured and explained by a computational cognitive architecture (Sun, 2002, 2009). That is, it should be captured and explained based on adequate built-in representation of basic motivational, metacognitive, action decision making, reasoning, and other processes, within a (relatively) generic and comprehensive computational model of the mind. Such representations and related mechanisms and processes may capture the interaction of internally felt needs and external environmental factors in determining goals and actions by individuals. They may capture the relative invariance within an individual in terms of behavioral propensities and inclinations at different times and with regard to different situations (social or physical), as well as behavioral variability. The (relative) invariance of personality has been extensively argued for in personality psychology (e.g., Caprara & Cervone, 2000; Epstein, 1982; Hall & Gardner, 1985; Maddi, 1996; Ryckmann, 1993; Staub, 1980; Wiggins, 1996); it can be captured computationally.

An outline of our general framework can be summarized as follows (cf. Cervone, 2004; Mayer, 2005; Read et al., 2010; Winter, John, Stewart, Klohnen, & Duncan, 1998): Within a comprehensive cognitive architecture, there is constant interaction among its subsystems: the motivational subsystem, the metacognitive subsystem, the action-centered (procedural) subsystem, and the non-action-centered (declarative) subsystem. Within the motivational subsystem, there is a set of basic motives, termed *primary drives*, which are universal across individuals. Individual differences may be explained (in part) by the differences across this set of drives in terms of drive strengths (activations) in different situations by different individuals. These drives, with their different strengths, lead to setting of different goals as well as major cognitive parameters by the metacognitive subsystem. Individual differences in terms of drive strengths are consequently reflected in the resulting goals as well as major cognitive parameters. On the basis of the goals set and the cognitive parameters chosen, an individual makes action decisions, within the action-centered subsystem, possibly in consultation with the non-action-centered subsystem using its declarative knowledge. Thus their actions reflect their fundamental individual differences as well as situational factors as a result. Their actions in turn affect the world in which they act.

As such, personality results from the complex interaction of a number of psychological entities, mechanisms, and processes, as well as their interaction with the world. Hence, computational simulations of personality based on a cognitive architecture can be useful in supplementing alternative approaches: Computational modeling and simulations may enable us to see exactly how these entities, mechanisms, and processes interact.

Why are all these mechanisms and processes needed to account for human personality? It is conceivable that these individual aspects or components of the mind exert a considerable influence individually or collectively on the personality of an individual. For example, Sun and Wilson (2011) and Deci (1980) argued that motivational representations are important for capturing personality. Similarly, behaviors (actions) are important, at least for measuring personality (Sun, 2002). Existing computational models of personality tend to be specialized, missing, for example, well-developed models of motivation, reasoning, metacognition, and so on. They tend to ignore some essential psychological distinctions (e.g., implicit vs. explicit; more later). They often rely on extensive parameter tweaking to generate desired outcomes. The upshot is that a model of personality should be a model of cognitive architecture (or a model of personhood, more generally), including all of these aspects enumerated earlier (and possibly more).

This is not to say that other aspects of personhood, besides a cognitive architecture, such as various aspects of the biological or social being, are not important. A model of personality should ideally be a model of personhood, biological, psychological, and social. The fact, however, is that there is currently no comprehensive model of the physical and physiological person available, beside relatively comprehensive models of the mind, that is, computational cognitive architectures. Given that we currently only have cognitive architectures available, which presumably capture some of the most important parts of the personhood, it seems wise to start with cognitive architectures. (Sociocultural aspects are dealt with elsewhere; see Sun, 2006, 2012; Wilson & Sun, in preparation.)

Thus, a variety of computational simulations of personality can be performed through a cognitive architecture as sketched above. The simulations should naturally fall out of the cognitive architecture, without any significant alterations or additions. They may help to clarify issues of personality in an exact, mechanistic, and process-based way. Practically speaking, the resulting personality model may be applied in practically useful ways. In particular, a model of personality may be important for simulating some social phenomena; that is, it may be applied to cognitive social simulation (Sun, 2006, 2012), which is one of our goals for the future.