Dual-Process Theories and Cognitive Architectures ·

Ron Sun
Rensselaer Polytechnic Institute, Troy, NY, USA
rsun@rpi.edu

Abstract

The distinction between intuitive and reflective thinking is arguably one of the most important distinctions in cognitive science. This talk will explore implicit processes (intuitive thinking) in a variety of tasks and domains, and their interaction with explicit processes (reflective thinking). A cognitive architecture will be used to addresses, in a mechanistic and process sense, such issues, including the relation, interaction, and competition between implicit and explicit processes. Such cognitive-psychological theories have serious implications for developing hybrid neural-symbolic models.

1 Introduction

The distinction between *intuitive* and *reflective* thinking has been, arguably, one of the most important distinctions in cognitive science. There are many dual-process theories (two-system views) out there, as they seem to have captured popular imagination nowadays. However, although the distinction itself is important, these terms have been somewhat ambiguous. Not much finer-grained analysis has been done, especially not in a precise, mechanistic, process-based way. In this article, towards developing a more fine-grained and comprehensive framework, I will adopt the somewhat better terms of implicit and explicit processes, and present a more nuanced view (centered on "cognitive architecture").

Given that there have already been an overwhelming amount of research on explicit processes ("reflective thinking"), it is important, in studying the human mind, to emphasize implicit processes (including intuition). I would argue that we need to treat them as an integral part of human thinking, reasoning, and decision-making, not as an add-on. Therefore, we need to explore implicit processes in a variety of tasks and domains, and their interaction with explicit processes (reflective thinking). A theoretical model will be presented that addresses, in a mechanistic, process-based way, such issues. Issues addressed will include different types of implicit processes, their relation to explicit processes, and their relative speeds in relation to explicit processes.

2 Some Background

There are many dual-process theories (two-system views) available. One such view was proposed early on in Sun (1994, 1995). In Sun (1994), the two systems were characterized as follows:

> "It is assumed in this work that cognitive processes are carried out in two distinct `levels' with qualitatively different mechanisms. Each level encodes a fairly complete set of knowledge for its processing, and the coverage of the two sets of knowledge encoded by the two levels overlaps substantially." (Sun, 1994)

That is, the two "levels" (i.e., two modules or two components) encode somewhat similar content. But they encode their content in different ways: Symbolic versus subsybmolic representations were used, respectively. Symbolic representation is used by explicit processes at one "level", and subsymbolic representation is used by implicit processes at another. Therefore, different mechanisms are involved at these two "levels". It was hypothesized in Sun (1994) that these two different "levels" can potentially work together synergistically, complementing and supplementing each other. This is, in part, the reason why there are these two levels.

However, some more recent two-system views are more contentious. For instance, a more recent two-system view (dual-process theory) was proposed by Kahneman (2003). The gist of his ideas was as follows: There are two styles of processing: intuition and reasoning. Intuition (or System 1) is based on associative reasoning, fast and automatic, involving strong emotional bonds, based on formed habits, and difficult to change or manipulate. Reasoning (or System 2) is slower, more volatile, and subject to conscious judgments and attitudes.

Evans (2003) espoused this view. According to him, System 1 is "rapid, parallel and automatic in nature: only their final product is posted in consciousness"; he also notes the "domain-specific nature of the learning". System 2 is "slow and sequential in nature and makes use of the central working memory system", and "permits abstract hypothetical thinking that cannot be achieved by System 1"

But such claims seem simplistic. For one thing, intuition can be very slow (Helie and Sun, 2010; Bowers et al., 1990). For another, intuition can be subject to conscious control and manipulation; that is, it may not be entirely "automatic" (Berry, 1991; Curran and Keele, 1993; Stadler, 1995). Furthermore, intuition can be subject to conscious "judgment" (Libet, 1987; Gathercole, 2003). And so on.

To come up with more nuanced and more detailed characterization, it is important that we ask some key questions. For instance, for either type of processes, there can be the following relevant questions:

- How deep is the processing (in terms of precision, certainty, and so on)?
- How much information is involved (how broad is the processing)?
- How incomplete, inconsistent, or uncertain is the information available?
- How many processing cycles are needed considering the factors above?

And many other similar or related questions.

3 A Theoretical Framework

In order to sort out these issues and questions, below, I will present a theoretical framework that can potentially provide some clarity to these issues and questions. The framework is based on the CLARION cognitive architecture (Sun, 2002, 2003, 2014), viewed at a theoretical level, as a conceptual tool for theoretical interpretations and explanations (Sun, 2009).

The theoretical framework consists of a number of basic principles. The first major point in this theoretical framework is the division between procedural and declarative processes, which is rather uncontroversial (Anderson and Lebiere, 1998; Tulving, 1985). The next two points concerns the division between implicit and explicit processes. They are unique to this theoretical framework (and thus may require some justifications; see, e.g., Sun, 2012, 2014). The second major point is the division between implicit and explicit procedural processes (Sun et al., 2005). The third major point is the division between implicit and explicit declarative processes (Helie and Sun, 2010). Therefore, in this framework, there is a four-way division: implicit and explicit procedural processes and implicit and explicit declarative processes.

The divisions above may be related to some existing computational paradigms, for example, symbolic-localist versus connectionist distributed representation (Sun, 1994, 1995). As has been extensively argued before (Sun, 1994, 2002), the relatively inaccessible nature of implicit knowledge may be captured by distributed representation (Rumelhart et al., 1986), because distributed representational units are subsymbolic and generally not individually meaningful. This characteristic of distributed representation, which renders the representational form less accessible computationally, accords well with the relative inaccessibility of implicit knowledge in a phenomenological sense. In contrast, explicit knowledge may be captured by symbolic or localist representation, in which each unit is more easily interpretable and has a clearer conceptual meaning.

4. A Sketch of A Cognitive Architecture

Now that the basic principles have been enumerated, I will sketch an overall picture of the CLARION cognitive architecture itself, which is centered on these principles (without getting into too much technical details though).

CLARION is a generic "cognitive architecture"---a comprehensive model of psychological processes of a wide variety, specified computationally. It has been described in detail and justified on the basis of psychological data (Sun, 2002, 2003, 2014). CLARION consists of a number of subsystems. Its subsystems include the Action-Centered Subsystem (the ACS), the Non-Action-Centered Subsystem (the NACS), the Motivational Subsystem (the MS), and the Meta-Cognitive Subsystem (the MCS). Each of these subsystems consists of two "levels" of representations, mechanisms, and processes as theoretically posited earlier (see also Sun, 2002). Generally speaking, in each subsystem, the "top level" encodes explicit knowledge (using symbolic-localist representations) and the "bottom level" encodes implicit knowledge (using distributed representations; Rumelhart et al., 1986).

Among these subsystems, the Action-Centered Subsystem is responsible for procedural processes, that is, to control actions (regardless of whether the actions are for external physical movements or for internal mental operations) utilizing and maintaining procedural knowledge. Among procedural processes, implicit procedural processes are captured by MLP (i.e., Backpropagation networks; at the bottom level of the ACS within the cognitive architecture). Explicit procedural processes, on the other hand, are captured by explicit "action rules" (at the top level of the ACS).

The Non-Action-Centered Subsystem is responsible for declarative processes, that is, to maintain and utilize declarative (non-action-centered) knowledge for information and inferences. Among these processes, implicit declarative processes are captured by associative memory networks (Hopfield type networks or MLP). Explicit declarative processes are captured by explicit "associative rules".

The Motivational Subsystem is responsible for motivational dynamics, that is, to provide underlying motivations for perception, action, and cognition (in terms of providing impetus and feedback). Implicit motivational processes are comprised of drive activations, captured by MLP. Explicit motivational processes are centered on explicit goals.

The Meta-cognitive Subsystem is responsible for metacognitive functions; that is, its responsibility is to monitor, direct, and modify the operations of the other subsystems dynamically. Implicit metacognitive processes are captured by MLP, while explicit metacognitive processes are captured by explicit rules.

The two levels within each subsystem interact, for example, by cooperating in action decision-making within the ACS, through integration of the action recommendations from the two levels of the ACS, as well as by cooperating in learning (more later; Sun, 2002). See Figure 1 for a sketch of the CLARION cognitive architecture**.**

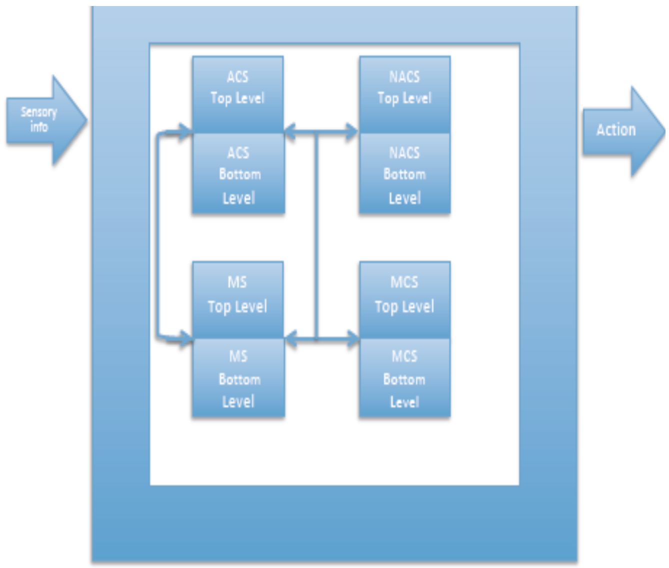

Figure 1. The four subsystems of CLARION.

5. Interpreting Psychological Notions

Based on the framework above (i.e., the CLARION cognitive architecture), we may re-interpret some folk psychological notions, to hopefully give them some clarity.

For instance, the notion of *instinct* may be made more precise by appealing to the framework of a cognitive architecture. Instinct involves mostly implicit processes and is mostly concerned with action. Within CLARION, instinct may be roughly equated with the following chain of activation: stimuli → drive → goal → action. This chain goes from stimuli received to the MS, the MCS, and eventually the ACS. That is, stimuli activate drives (especially those representing innate motives), drive activations lead to goal setting in a direct, implicit way (mostly innate), and based on the goal set, actions are selected in an implicit way to achieve the goal. Instinct is mostly implicit, but it may become more explicit, especially with regard to the part of "goal → action" (Sun et al., 2001).

For another instance, the notion of *intuition* can also be made more precise by using the CLARION framework. Intuition, according to CLARION, is roughly the following chain: stimuli → drive → goal → implicit reasoning. This chain goes from stimuli received to the MS, the MCS, and the NACS. As such, intuition mostly involves implicit declarative processes within the NACS, including the functionality of associative memory retrieval, soft constraint satisfaction, and partial pattern completion. Intuitive processes are often complementary to explicit reasoning, and the two types are used often in conjunction with each other (Helie and Sun, 2010).

Some other folk psychological notions may be re-interpreted and made more precise in a similar manner. For example, the notion of *creativity* may be captured within the CLARION framework. Creativity may be achieved through complex, multi-phased implicit-explicit interaction, that is, through the interplay between intuition and explicit reasoning, according to Helie and Sun's (2010) theory of creative problem solving---a theory derived from the CLARION cognitive architecture. It is a 3-phase model, which includes (1) the explicit phase: process given information; (2) the implicit phase: develop intuition using implicit declarative knowledge; finally the intuition emerge into explicit processes and therefore (3) the explicit phase: verify and validate the result using explicit declarative knowledge. See Helie and Sun (2010) for further details. This theory has been successful in accounting for a variety of empirical data.

What about the competition among these different types of processes, especially in terms of their relative speeds (time courses) as alluded to earlier? There was a question raised earlier concerning fast versus slow processes with regard to different dual-process theories (two-systems views). The twin divisions in CLARION, procedural versus declarative and implicit versus explicit, definitely have implications for identifying slow versus fast processes across the systems (components). For instance, we may question conventional wisdom on a number of issues in this regard (instead of simply assuming the seemingly obvious as in some of the existing views/theories):

- In terms of the division between procedural and declarative processes, can fast procedural versus slow declarative processes be posited?
- In terms of the division between implicit and explicit procedural processes, can fast implicit versus slow explicit processes be posited?
- In terms of the division between implicit and explicit declarative processes, can fast implicit versus slow explicit processes be likewise posited?
- What about relative speeds if we consider the four-way division together?

And so on. These conjectures implied by the questions above may be true to some extent, but not exactly accurate. The whole picture is not so simple.

In this regard, we may view existing models and simulations of these different types of processes as a form of theoretical interpretation concerning their time courses. In that case, we have the following potential answers:

- Fast procedural versus slow declarative: Fortunately, this is generally true if we examine many existing models and simulations (Anderson and Lebiere, 1998; Sun, 2003, 2014).
- Fast implicit versus slow explicit procedural processes: It is, again, generally true, using theoretical interpretations through modeling and simulation (Sun et al., 2001, 2005).

- Fast implicit versus slow explicit declarative processes: This, however, is generally not true. Intuition (implicit declarative processes) may (or may not) take a long time, compare with explicit declarative processes. See, for example, Helie and Sun (2010) and Bowers et al. (1990).

We need to be careful in making sweeping generalizations. Often, we need to characterize different types of processes in a more fine-grained fashion. Characteristics of different processes may also vary in relation to contexts such as task demands.

A number of empirical and simulation studies have been conducted within the framework of CLARION that shed light on these issues, and substantiate the points made above. See, for example, Sun et al. (2009), Sun and Zhang (2006), Helie and Sun (2010), and so on.

Dual-process theories (two-system views) may arguably have significant implications for neural symbolic processing and especially hybrid neural-symbolic systems. For one thing, they may serve as the theoretical basis and the justifications for some forms of hybrid neural-symbolic models that juxtapose symbolic and subsymbolic components. If cogniitive-psychological realism is what one wants to achieve in developing such models, dual-process theories must be taken into serious consideration, and be used as guides in developing such models. See, for example, Sun (2002, 2006, 2014).

Acknowledgments

The work overviewed here has been supported in part by the ONR grants N00014-08-1-0068 and N00014-13-1-0342. Thanks are due to my students and collaborators, past and present, whose work has been reviewed here. The CLARION software, along with simulation examples, may be found at: www.clarioncognitivearchitecture.com (courtesy of Nick Wilson and Mike Lynch).

References

J. R. Anderson and C. Lebiere, (1998). *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Mahwah, NJ.

D. Berry, (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, 43A, 881-906.

K. Bowers, G. Regehr, C. Balthazard, and Parker, (1990). Intuition in the context of discovery. *Cognitive Psychology*. 22. 72-110.

T. Curran and S. Keele, (1993). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 19, 189-202.

J. St.B.T. Evans, (2003). In two minds: dual-process accounts of reasoning, *Trends in Cognitive Sciences*, 7 (10), 454-459.

S. Gathercole, (2003). *Short-term and Working Memory*. Taylor and Francis, London.

S. Helie and R. Sun, (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, 117, 994-1024.

D. Kahneman (2003). A perspective on judgment and choice: mapping bounded rationality. Am Psychol. 58(9):697-720.

B. Libet, (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*. 8, 529-566.

D. Rumelhart, J. McClelland and the PDP Research Group, (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, MIT Press, Cambridge, MA.

M. A. Stadler, (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 674-685.

R. Sun, (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* John Wiley and Sons, New York.

R. Sun, (1995). Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence* (AIJ). Vol.75, No.2, pp.241-296.

R. Sun, (2002). *Duality of the Mind: A Bottom-up Approach Toward Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

R. Sun, (2003). *A Tutorial on CLARION 5.0.* Technical Report, Cognitive Science Department, Rensselaer Polytechnic Institute. http://www.cogsci.rpi.edu/rsun/sun.tutorial.pdf

R. Sun, (2006). *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge University Press, New York.

R. Sun, (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research*, 10 (2), 124-140.

R. Sun, (2014). *Anatomy of Mind*. Oxford University Press, New York. In press.

R. Sun, Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25*, 203-244.

R. Sun, Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112*, 159-192.

R. Sun and X. Zhang, (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, Vol.18, No.2, pp.169-191.

R. Sun, X. Zhang, and R. Mathews (2009). Capturing human data in a letter counting task: Accessibility and action-centeredness in representing cognitive skills . *Neural Networks*, Vol.22, pp.15-29.

R. Sun, X. Zhang, P. Slusarz, and R. Mathews, (2007). The interaction of implicit learning, explicit hypothesis testing learning, and implicit-to-explicit knowledge extraction. *Neural Networks*, Vol.20, No.1, pp.34-47.

E. Tulving, (1985). How many memory systems are there? *American Psychologist*, 40,385–398.