Oxford Handbooks Online

The CLARION Cognitive Architecture: Toward a Comprehensive Theory of the Mind

Ron Sun

The Oxford Handbook of Cognitive Science
Edited by Susan E. F. Chipman

Print Publication Date: Oct 2017 Subject: Psychology, Cognitive Psychology
Online Publication Date: May 2015 DOI: 10.1093/oxfordhb/9780199842193.013.11

Abstract and Keywords

This chapter presents a hybrid cognitive architecture CLARION, which is significantly different from most existing cognitive architectures in several important respects. CLARION is hybrid in that it (a) combines connectionist and symbolic representations, (b) combines implicit and explicit psychological processes, and (c) combines cognition (in the narrow sense) and other psychological processes. Overall, CLARION is a modularly structured cognitive architecture consisting of a number of functional subsystems. It also has a dual representational structure, with both implicit and explicit representations. CLARION has been successful in capturing a variety of psychological processes in a variety of task domains based on its structuring of functional modules.

Keywords: cognitive architecture, cognitive modeling, CLARION, implicit processes, motivation, meta-cognition

Introduction

In this chapter, I discuss a hybrid cognitive architecture CLARION that is significantly different from most existing cognitive architectures in several important respects. For one thing, the CLARION cognitive architecture is hybrid in that it (a) combines connectionist and symbolic representations computationally, (b) combines implicit and explicit psychological processes, and (c) combines cognition (in the narrow sense) and other psychological processes (such as motivation and emotion).

In presenting an overview of this cognitive architecture, I start with a look at some general ideas underlying the development of the cognitive architecture. CLARION represents an attempt at tackling a host of issues arising from computational cognitive modeling that are not adequately addressed by many existing cognitive architectures (Sun, 2002). Overall, CLARION, a modularly structured cognitive architecture, is made up of a number of functional subsystems (and many modules within them). It also has a dual representational structure, with both implicit and explicit representations (in separate modules within each subsystem). CLARION has been successful in capturing a variety of psychological processes in a variety of task domains based on its unique structuring of functional modules (as well as its internal processing within and across modules; Sun, 2002, 2003).

An important assumption of CLARION, one that has been argued for time and again (see, e.g., Sun, 2002), is the dichotomy of implicit and explicit processes. Generally speaking, implicit processes are less accessible and more “holistic,” whereas explicit processes are more accessible and more crisp (Reber, 1989; Sun, 2002). This dichotomy is closely related to some other well-known dichotomies in cognitive science, for example, the dichotomy of symbolic versus subsymbolic processing (Sun, 1995). The dichotomy can be justified psychologically by the (p. 118) voluminous empirical studies of implicit and explicit learning, implicit and explicit memory, implicit and explicit perception, and so on (Cleeremans et al., 1998; Reber, 1989; Seger, 1994). See Sun (2002) for an extensive treatment of this distinction.

In addition to this characteristic, a number of other CLARION characteristics are also important. For instance, one particularly important characteristic of CLARION is its focus on cognition-motivation-environment interaction. The essential motivations of an agent—its biological needs in particular—arise naturally, prior to cognition (but interact with cognition). In a way, cognition has evolved to serve the essential needs of an agent. Cognition, in the process of helping to satisfy needs and following motivational forces, has to take into account environments, their regularities and structures. Thus, cognition bridges the needs and motivations of an agent and its environments (be it physical or social), thereby linking all three in a triad. In this regard, CLARION includes drives and goals for capturing human motivation (more on this later).

Yet another important characteristic of CLARION is that an agent may learn on its own, regardless of whether there is a priori or externally provided domain-specific knowledge. Learning may proceed autonomously on a trial-and-error basis. Furthermore, on the basis of implicit knowledge acquired via trial-and-error learning, through a bootstrapping process (or “bottom-up learning,” as it has been termed in Sun et al., 2001), explicit knowledge may be developed, also in a gradual and incremental fashion. This is different from many other existing cognitive architectures (e.g., Anderson and Lebiere, 1998).

It should be noted that, although CLARION addresses autonomous trial-and-error and bottom-up learning, it can also capture innate biases and innate behavioral propensities. Innate biases and propensities may interact with trial-and-error and bottom-up learning by way of constraining, guiding, and facilitating learning. Opposite of bottom-up learning, top-down learning—that is, assimilation of explicit knowledge from external sources—is also captured in CLARION.

In the remainder of this chapter, first, a brief description of CLARION will be given, including its theoretical underpinnings and empirical support. Then, a number of examples of using CLARION to account for and explain psychological processes will be provided. With regard to examples of the psychological phenomena that CLARION can capture and explain, focus is on skill acquisition and categorical reasoning, as well as on motivation, emotion, and personality. In the last section, a discussion of different views and a comparison to other models will be given.

A Sketch of the Architecture

Overview

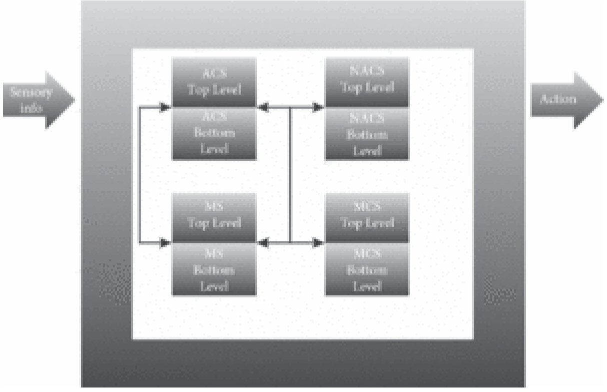
Overall, CLARION is an integrative cognitive architecture consisting of a number of subsystems, with a dual representational memory structure in each subsystem (with implicit versus explicit representations; see Cleeremans et al., 1998; Reber, 1989; Seger, 1994; Sun et al., 2005). CLARION is intended for capturing all the essential psychological processes within an individual in accordance with an ecological-functional perspective (Sun, 2002). It therefore contains all these distinct subsystems that are necessary for psychological functioning.

Accordingly, its subsystems include the action-centered subsystem (ACS), the non-action-centered subsystem (NACS), the motivational subsystem (MS), and the meta-cognitive subsystem (MCS). Their respective roles are as follows:

- The role of the ACS is to control actions (based on procedural knowledge), regardless of whether the actions are for external physical movements or internal mental operations.
- The role of the NACS is to maintain general (i.e., declarative) knowledge for retrieval of appropriate information and inferences on that basis (ultimately in the service of action decision making by the ACS).
- The role of the MS is to provide underlying motivations for perception, action, and cognition.
- The role of the MCS is to monitor, direct, and modify the operations of the other subsystems for the sake of performance.

Each of these subsystems serves a unique function, and together they form a functioning cognitive architecture (see Figure 6.1).

Each of these subsystems consists of two “levels” of representations (i.e., two sets of modules or components with two different kinds of representations interacting with each other), which together constitute a dual representational structure. Generally, in each subsystem, the top level encodes explicit knowledge (explicit memory) and the bottom level encodes implicit knowledge (implicit memory). This implicit-explicit distinction has been argued for amply (more details on this later). (p. 119)



Click to view larger

Fig. 6.1 The four subsystems of the CLARION cognitive architecture. The major information flows are shown with arrows.

The relatively inaccessible nature of implicit knowledge in implicit memory may be captured by subsymbolic, distributed representation provided, for example, by a back-propagation network (Rumelhart et al., 1986). This is because distributed representational units in the hidden layer(s) of a back-propagation network are capable of accomplishing computations but are subsymbolic and generally not individually meaningful (Rumelhart et al., 1986; Sun, 1995). This characteristic of distributed representation, which renders the representational form less accessible, accords well with the relative inaccessibility of implicit knowledge (Reber, 1989). In contrast, explicit knowledge in explicit memory may be captured by symbolic or localist representation, in which each unit is more easily interpretable and has a clearer conceptual meaning. This characteristic of symbolic or localist representation captures the characteristic of explicit knowledge as being more accessible and more manipulable (Sun, 1995). (Accessibility here refers to the direct and immediate availability of mental content for the major operations that are responsible for or concomitant with consciousness, such as introspection and verbal reporting.)

The dichotomous difference between the representations of these two different types leads naturally to a two-level structure within each subsystem, whereby each level uses one kind of representation and captures one corresponding type of process (implicit or explicit).

There have been some indications that one level (the explicit level) is evolutionarily newer than the other (see, e.g., LeDoux, 1996). In addition to this difference, the juxtaposition of the two levels is also functional. This is at least in part because the interaction between the two levels may lead to synergy in the form of better performance in a variety of circumstances (see Sun et al., 2005, for details). Therefore, such a division between the two levels may conceivably be favored by natural selection.

The Action-Centered and Non-Action-Centered Subsystems

The ACS controls actions based on procedural knowledge in both implicit and explicit forms (at the bottom and top levels of the ACS, respectively). The NACS deals with declarative knowledge in both implicit and explicit forms (at the bottom and top levels of the NACS, respectively). Such a division may be justified as follows.

The distinction between implicit and explicit processes/memory needs to be examined first. The theoretical distinction between implicit and explicit processes, as well as its significance, has been variously argued for in many theories (e.g., Reber, 1989; Sun, 1995, 2002). The distinction between implicit and explicit memory has been empirically demonstrated in the implicit memory literature (Roediger, 1990; Schacter, 1987). Work on amnesics showed that they might have intact implicit memory while (p. 120) their explicit memory was severely impaired. Jacoby (e.g., 1983) demonstrated that implicit and explicit measures might be dissociated among normal subjects as well. Toth, Reingold, and Jacoby (1994) devised the inclusion-exclusion procedure, which provided strong indications of the dissociation.

The distinction has also been empirically demonstrated in the implicit learning literature (e.g., Cleeremans et al., 1998; Reber, 1989; Seger, 1994). For example, serial reaction time tasks probe learning of a repeating sequence. It was found that there was a significant reduction in response time to repeating sequences (compared to random sequences). However, participants might not be able to explicitly report the repeating sequence and were often unaware that a repeating sequence was involved (see, e.g., Lewicki et al., 1987).

There are many other tasks that are similar in this regard, such as various concept learning, reasoning, automatization, and instrumental conditioning tasks (see Sun, 2002, for a review; see also Evans and Frankish, 2009). Together, they demonstrated the distinction between implicit and explicit processes. Although some researchers have disputed the existence of implicit processes based on the imperfection and incompleteness of tests for explicit knowledge, there is an overwhelming amount of evidence in support of the distinction (Sun et al., 2005).

A further question is whether these different types of processes/memory reside in separate modules. There have been debates on this question. Sun (2002) and Sun et al. (2005) provided some theoretical interpretations of existing learning data based on the multiple modules view. In social psychology, there have been a number of models developed that are roughly based on the coexistence of implicit and explicit modules (Evans & Frankish, 2009). This division of modules is functional: The separation of the two types of processes enables the application of each type separately, as appropriate for different situations (see Sun & Mathews, 2005). Furthermore, the interaction of the two separate types may lead to overall better performance (i.e., synergy between the two types) under proper circumstances (as demonstrated in, e.g., Sun et al., 2005).

I now turn to the distinction between procedural and declarative processes (i.e., action-centered and non-action-centered processes in the ACS and the NACS of CLARION, respectively) and its orthogonality with the implicit-explicit distinction. Procedural memory contains knowledge that is specifically concerned with actions in various circumstances, that is, how to do things. Declarative memory contains knowledge that is not specifically concerned with actions but is more about objects, events, and so on in generic terms (i.e., the “what,” not the “how”). Evidence in support of this distinction includes voluminous studies of skill acquisition in both high- and low-level skill domains (e.g., Anderson & Lebiere, 1998; Ackerman & Kanfer, 2004). These studies showed that making this distinction provided useful insight in interpreting a range of data and phenomena.

The relation between the procedural-declarative distinction and the implicit-explicit distinction, however, needs to be examined. Often, declarative knowledge is assumed to be consciously accessible (i.e., explicit) whereas procedural knowledge is not. Thus, the two dichotomies are merged into one. Alternatively, it is often assumed that each individual piece of knowledge, be it procedural or declarative, involves both subsymbolic and symbolic representation. One interpretation is that the symbolic representation is explicit, whereas the subsymbolic representation is implicit (either for declarative knowledge or for both declarative and procedural knowledge).

According to the first view, the difference in action-centeredness seems the main factor in distinguishing the two types of knowledge, whereas accessibility (i.e., implicitness versus explicitness) is a secondary factor. This view confounds two aspects—action-centeredness and accessibility—and can be made clearer by separating the two dimensions. There are reasons to believe that action-centeredness does not necessarily go with implicitness (inaccessibility), as shown, for example, by the experiments of Stanley, Mathews, Buss, and Kotler-Cope (1989) or Sun et al. (2001). Likewise, non-action-centeredness does not necessarily go with explicitness (accessibility) either, as shown by conceptual priming and other implicit memory experiments (e.g., Moscovitch & Umilta, 1991; Schacter, 1987) or by experiments demonstrating implicit statistical information (Hasher & Zacks, 1979; Nisbett & Wilson, 1977). Some might group all implicit memory (including semantic, associative, and conceptual priming) under procedural memory (e.g., Squire, 1987), but such views confound the definition of "procedural" and thus are not adopted here.

The alternative view, that each individual piece of knowledge (either procedural, declarative, or both) involves both implicit and explicit parts, is also problematic. The underlying assumption that every (p. 121) piece of knowledge (either declarative, procedural, or both) has an explicit part contradicts the fact that some knowledge may be completely implicit (Cleeremans et al., 1998; Lewicki et al., 1987). This contradiction raises the question of whether a tight coupling or a more separate organization (i.e., one having these two types in separate modules) makes better sense.

As an alternative to these views, CLARION posits the separation of the two dichotomies (Sun, Zhang, & Mathews, 2009). Based on empirical data, Willingham (1998) argued that motor skills (a kind of procedural knowledge) consisted of both implicit and explicit processes. Rosenbaum et al. (2001) argued based on empirical data that both intellectual skills and perceptual-motor skills were made up of implicit and explicit knowledge. In other words, procedural knowledge (action-centered knowledge), in situations ranging from high-level intellectual skills to perceptual-motor skills, may be divided into implicit and explicit procedural memory.

Similarly, declarative knowledge (non-action-centered knowledge) may also be divided into implicit and explicit memory (Tulving & Schacter, 1990). In terms of functional consideration, having separate implicit and explicit declarative memory allows different tasks to be tackled simultaneously (e.g., while thinking explicitly about one task, one can allow intuition to work on another). Sun and Zhang (2006) showed that, through dividing declarative memory into explicit and implicit modules, similarity-based reasoning (SBR) data could be naturally accounted for (through the interaction of the two types). Furthermore, Helie and Sun (2010) showed that this division accounted well for creative problem solving (through the interaction of the two), which otherwise would be difficult to account for.

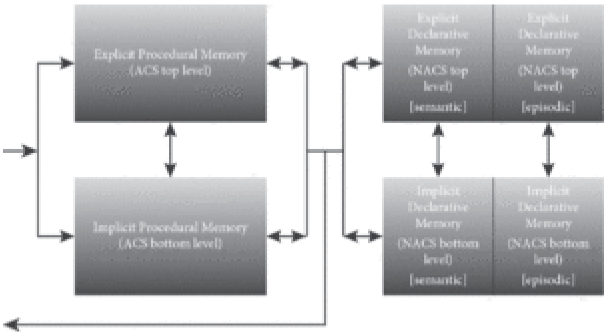
In CLARION, procedural and declarative knowledge reside in procedural and declarative memory (i.e., the ACS and the NACS), respectively. Procedural knowledge (in the ACS) is represented by either action rules (explicit) or action neural networks (implicit), both of which are centered on situation-action mappings. Declarative knowledge (in the NACS), on the other hand, is represented by either associative rules (explicit) or associative neural networks (implicit), in both of which knowledge is represented in a non-action-centered way.

As mentioned earlier, in a similar fashion but orthogonally, implicitness/explicitness is also distinguished based on representation in CLARION. Implicit memory is represented using connectionist distributed representation (such as in the hidden layer of a back-propagation network), which is less accessible (Sun, 2002), whereas explicit memory is represented using symbolic/localist representation, which is relatively more accessible.

This four-way division is functional, because of (1) the division of labor between explicit and implicit memory (while one is used for storing explicit information that is more crisp, the other is used for storing implicit information that is more complex) and (2) the division of labor between declarative and procedural memory (while one is used for storing general information, the other is used for storing information oriented specifically toward action decision making). The divisions of labor led to both the separation and the interaction of these different types. The separation ensures that different types of information may be found separately and thus relatively easily, whereas the interaction among different types helps to bring together different types of information when needed (Klein et al., 2002; Sun et al., 2007, 2009) to ensure better performance and synergy, as mentioned earlier (Sun et al., 2005). Furthermore, the separation makes it possible for different memories to work on different tasks simultaneously and thus enhances the overall functionality.

However, within the declarative memory (i.e., within the NACS of CLARION), the distinction between episodic and semantic memory and its orthogonality with the implicit-explicit distinction need to be mentioned briefly also. This point has been argued in detail in Sun (2012). The distinction between episodic and semantic memory has been developed by Quillian (1968) and many others. Rogers (2008) and Norman et al. (2008) discussed various models of semantic and episodic memory, respectively. Furthermore, the distinction between implicit and explicit semantic memory is supported by data discussed by Schacter (1987), Roediger (1990), and others. Sun and Zhang (2006) showed how the division of implicit and explicit semantic memory might account for categorical inferences; Helie and Sun (2010) showed how the division of implicit and explicit semantic memory might explain creative problem solving. Similarly, the distinction between implicit and explicit episodic memory has also been argued (Sun, 2012).

The overall structure of the ACS and the NACS in CLARION is shown in Figure 6.2. A combination of various learning methods, including reinforcement, associative, and hypothesis (p. 122) testing learning, is used to acquire these types of knowledge (Sun, 2002, 2003). The ACS controls all types of actions and procedural knowledge within the ACS may be learned through trial-and-error interactions with the world. Such learning, using reinforcement learning algorithms, leads to implicit procedural knowledge (at the bottom level of the ACS), which may in turn lead to explicit procedural knowledge at the top level of the ACS (using bottom-up learning that turns implicit knowledge into explicit knowledge). There are also other ways in which explicit procedural knowledge may be learned (Sun, 2002). The two kinds of action recommendations (from the two levels of the ACS) may be integrated in some way, and a combined action recommendation is then produced that directs the action of the agent. On the other hand, for learning declarative knowledge within the NACS, explicit information from external sources may be utilized, and then the acquired explicit declarative knowledge (at the top level of the NACS) may be assimilated into implicit declarative knowledge (at the bottom level of the NACS). Declarative knowledge may also be the result of transfer from procedural learning. Experiences with perception, action, and decision making populate not only episodic memory in the NACS, but also the semantic memory within the NACS. For technical details, the reader is referred to Sun (2002, 2003) as well as Sun et al. (2009) and Helie and Sun (2010).


Explicit Procedural Memory (ACS top level)
Implicit Procedural Memory (ACS bottom level)
Explicit Declarative Memory (NACS top level) [semantic]
Explicit Declarative Memory (NACS top level) [episodic]
Implicit Declarative Memory (NACS bottom level) [semantic]
Implicit Declarative Memory (NACS bottom level) [episodic]

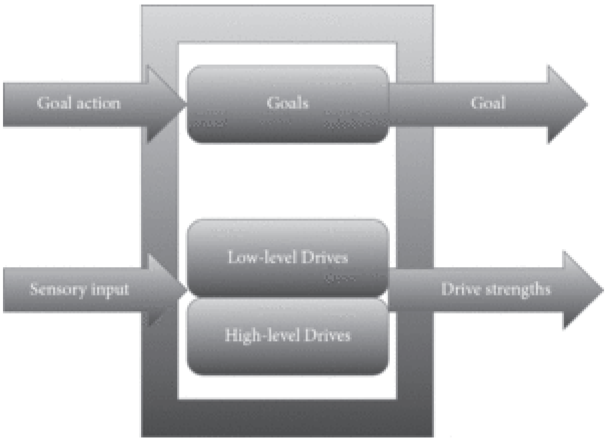
Click to view larger

Fig. 6.2 The essential memory modules. The leftmost lines show the input information to and output actions from the action-centered subsystem (ACS). The lines between the modules show the information flows (the working memory, goal structure, and sensory information store are used to facilitate the flows but are omitted for the sake of clarity). See Sun (2003) for technical details.

The Motivational and Meta-Cognitive Subsystems

Aside from the ACS and the NACS, the MS of CLARION is pertinent for addressing the cognition-motivation interaction: it is concerned with why an agent does what it does. Simply saying that an agent chooses actions within the ACS to maximize rewards or reinforcement leaves open the question of what determines reward or reinforcement. The relevance of the MS to the ACS lies in the fact that it provides the context in which the goal and reinforcement of the ACS are determined. It thereby influences the working of the ACS (and, by extension, the working of the NACS).

A dual motivational representation is in place in the MS (Sun, 2009). The explicit goals (such as “finding food”), which are essential to the working of the ACS as explained earlier, may be generated based on internal activation of drives (e.g., “being hungry”). The explicit representation of goals derives from and hinges upon implicit drives (see Figure 6.3). Sun (2009) presented theoretical justifications for this dual representation of motivation. Goals are different from drives: for example, (1) there may be multiple drives being activated at the same time, whereas there is usually only one goal being pursued at a time (although a goal may encode multiple action objectives; Sun, 2003). (2) Drives are often more diffused in focus, whereas goals are often more specific (McFarland, 1989). (3) Drives are more implicit, whereas goals are more (p. 123) explicit (Hull, 1951; Maslow, 1943; Murray, 1938). (4) Drives are often hardwired, whereas goals are more flexibly created, set, and carried out (Hull, 1951; Sun, 2009).



Click to view larger

Fig. 6.3 The structure of the motivational subsystem.

Specifically, we refer to as “primary drives” those motives that are essential and likely built-in (hardwired) to a significant extent to begin with. Some sample low-level primary drives (concerning mostly physiological needs) include food, water, reproduction, and so on (McDougall, 1936; Murray, 1938). Beyond such low-level drives are high-level drives, which are more socially oriented. Some of them are primary, in the sense of being more or less “hardwired” (e.g., dominance and power, fairness, and so on).[1] The primary drives in the MS of CLARION (low-level and high-level together) may be roughly defined as in Table 6.1 (see Sun, 2009, for further details).

This set of primary drives has been explored and justified in detail (e.g., Sun, 2003, 2009). Briefly, this set of hypothesized primary drives is essentially the same as Murray’s (1938), with only a few major differences. Similarly, comparing this set of hypothesized primary drives with Reiss (2004), one can see that they are highly similar (but with some differences). In addition, Schwartz’s (1994) 10 universal values bear some resemblance to the primary drives identified here, and each of his values can be derived from some primary drive or some combination of these primary drives. So, the prior work by these researchers in justifying their frameworks may be applied, to a significant extent, to this set of hypothesized primary drives as well (Maslow, 1987; McDougall, 1936; Murray, 1938; Reiss, 2004).

These drives may also be divided up into an approach system and an avoidance system. Gray and McNaughton (2000) argued that underlying extroversion was a behavioral approach system (BAS) and underlying neuroticism was a behavioral inhibition system (BIS). Others (e.g., Cacioppo, Gardner, & Berntson, 1999) also argued for similar distinctions between approach and avoidance systems. The approach system is sensitive to cues signaling rewards and results in active approach. The avoidance system is sensitive to cues of punishment or threat and results in avoidance of threatening situations, characterized by anxiety or fear (see Table 6.2).

Technically, the activation (strength) of each of these drives is calculated essentially by the product of a drive-specific stimulus level (which measures the pertinence of the current situation to the drive) and a drive-specific deficit level (which measures the internal inclination to activate the drive). These activated drives then compete for goal setting through a probability distribution.

The existence of drives and the need for goal setting lead to the need for meta-cognitive control and regulation. In CLARION, the MCS is closely tied to the MS. Control and regulation may be in the forms of setting goals (which are then used by the ACS) on the basis of drives, interrupting and changing ongoing processes in the ACS and the NACS, setting essential parameters of the ACS and the NACS, and so on. Control and regulation are also carried out through setting reinforcement functions (for reinforcement learning in the ACS) on the basis of drives and goals. (p. 124)

Table 6.1 The List of Primary Drives in CLARION

• Food	The drive to consume nourishment.
• Water	The drive to consume fluid.
• Sleep	The drive to rest and/or sleep
• Reproduction	The drive to mate.
• Avoiding danger	The drive to avoid situations that have the potential to be or already are physically harmful.
• Avoiding unpleasant stimuli	The drive to avoid situations that are physically (or emotionally) uncomfortable or negative in nature.
• Affiliation and belongingness	The drive to associate with other individuals and to be part of social groups.
• Dominance and power	The drive to have power over other individuals or groups.
• Recognition and achievement	The drive to excel and be viewed as competent at something.

• Autonomy	The drive to resist control or influence by others.
• Deference	The drive to willingly follow and serve a person of a higher status of some kind.
• Similance	The drive to identify with other individuals, to imitate others, and to go along with their actions.
• Fairness	The drive to ensure that one treats others fairly and is treated fairly by others.
• Honor	The drive to follow social norms and codes of behavior and to avoid blame.
• Nurturance	The drive to care for or attend to the needs of others who are in need.
• Conservation	The drive to conserve, to preserve, to organize, or to structure (e.g., one’s environment).
• Curiosity	The drive to explore, to discover, and to gain new knowledge.

Table 6.2 Approach Versus Avoidance Drives

Approach Drives	**Avoidance Drives**	**Both**
Food	Sleep	Affiliation and belongingness
Water	Avoiding danger	Similance
Reproduction	Avoiding unpleasant stimuli	Deference
Nurturance	Honor	Autonomy
Curiosity	Conservation	Fairness
Dominance and power		
Recognition and achievement		

(p. 125) Structurally, this subsystem is subdivided into a number of functional modules, including:

- the goal setting module,
- the reinforcement function module,
- the processing mode selection module.
- the input filtering/selection module,
- the output filtering/selection module,
- the parameter setting module (for setting learning rates, temperatures, etc.), and so on.

For example, in order to select a new goal, the goal setting module of the MCS calculates goal strengths based on information from the MS (e.g., the drive strengths), as well as the current sensory input. Then, a new goal is selected on the basis of the goal strengths. For the arguments in support of goal setting on the basis of implicit motives (i.e., drives), see Tolman (1932) and Deci (1980). See also related empirical findings such as those by Elliot and Thrash (2002).

For another example, the reinforcement function module of the MCS produces an evaluation of the current input state in relation to the current goal and the currently active drives: how much it satisfies the current goal and/or the activated drives. This evaluation is used as reinforcement (for reinforcement learning in the ACS; see, e.g., Montague, 1999, and Sun et al., 2001, regarding reinforcement in human learning).

Capturing and Explaining Psychological Phenomena

CLARION has been successful in simulating, accounting for, and thereby explaining a wide variety of psychological phenomena. For example, a number of well-known skill learning tasks have been simulated using CLARION, which span the spectrum ranging from simple reactive skills to complex cognitive skills. Also, reasoning, social simulation, meta-cognitive, and motivational tasks have been accounted for. While accounting for various psychological tasks, CLARION provides explanations that shed new light on underlying processes.

CLARION has also been validated by new human experiments specifically geared toward exploring the implications of CLARION. See, for example, Domangue et al. (2004), Lane et al. (2008), Sallas et al. (2007), Sun and Mathews (2005), and so on for experimental details and findings.

Some examples are in order. Two of the following subsections address typical cognitive tasks (involving the ACS and the NACS, respectively), and the other subsection addresses some psychological processes beyond the narrow definition of cognition.

Synergy of Implicit and Explicit Processes in Learning

The example below addresses the synergy between implicit and explicit procedural processes within the ACS.

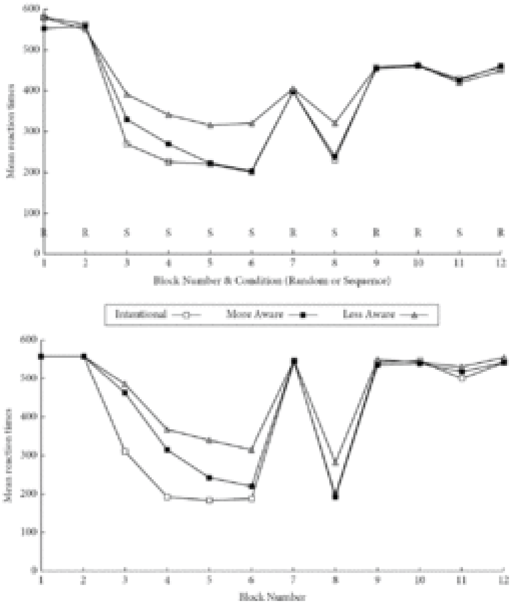
Task and Data

The task of Curran and Keele (1993) consisted of a repeating sequence of X marks, each appearing in one of four possible positions. Test participants were instructed to press the key corresponding to the position of each X mark. Reaction times of the participants were recorded. The experiment was divided into three phases (in succession): dual-task practice, single-task learning, and dual-task transfer. In the first phase, the positions of X marks were purely random to allow participants to get used to the task setting. This phase consisted of two blocks. Each block consisted of 120 trials. The second phase was when learning occurred: there were five sequence blocks (Blocks 3, 4, 5, 6, and 8) in which the positions followed a sequential pattern of length 6 (e.g., 1, 2, 3, 2, 4, 3) and one random block (Block 7). The third phase tested transfer to a dual-task condition (with a secondary tone counting task). Three random blocks (Blocks 9, 10, and 12) and a single sequence block (Block 11) were presented.

Three groups of participants were identified: “less aware,” “more aware,” and “intentional.” The intentional participants were given explicit instructions about the exact sequence used before learning started. The more aware participants were those who, after the experiment, correctly specified at least four out of six positions in the sequence used (which demonstrated their explicit knowledge); and the remaining participants were less aware subjects.

Statistical analysis showed that, although the intentional group performed the best, the more aware group performed close to the intentional group, and both performed significantly better than the less aware group. There was a significant interaction between group and block, demonstrating the effect of explicit knowledge. However, during the transfer to the dual-task condition (which was known to suppress mostly explicit knowledge), all three groups performed poorly, and there was no significant difference across the three groups. See Figure 6.4 for the reaction time data.

The finding of interest here is that the difference in explicit knowledge led to the difference in (p. 126) performance in phase 2, and the performance difference disappeared under the dual-task condition in phase 3.



Click to view larger

Fig. 6.4 Top: The reaction time data (in milliseconds) from Curran and Keele (1993). Bottom: The reaction time data (in milliseconds) from the CLARION simulation. R = random; S = sequence.

Simulation

The simulation of this task was carried out by the two levels of the ACS, which handled both the primary task and the secondary tone-counting task. As described in Sun et al. (2005), rule learning was carried out in a bottom-up fashion. The difference between less aware and more aware participants was captured by the difference in rule learning thresholds. To simulate less aware participants, higher thresholds were used so that they were less likely to develop explicit knowledge than more aware participants. To simulate the intentional group, the given sequence was coded as a set of a priori rules (before the training started).

To simulate the dual-task condition, the presentations of tones and lights were interleaved (as in Curran & Keele, 1993). Consistent with the existing understanding of the effect of dual tasks (which lay mostly in interfering with explicit processes; see Sun et al., 2001), it was hypothesized that the dual task interfered mostly with and thus reduced top-level activities. Thus, the rule learning thresholds were increased to reduce top-level activities (because rule learning was more effortful than rule application).

(p. 127) A linear transformation was used that turned error rate into reaction time. The results, as shown in Figure 6.4, captured the essential characteristics of the human data (a nonlinear transformation produced an even better fit; Sun et al., 2005).

In correspondence with the analysis of human data in this task, statistical analysis showed that, in the simulation data, there was a significant interaction between group and block, thus indicating a significant effect of explicit knowledge, similar to what was found in the human data. The more aware group and the intentional group performed significantly better than the less aware group, as in the human data. For the transfer to the dual-task condition, analysis showed that there was no significant difference among the three groups, showing the disappearance of the effect of explicit knowledge under the dual-task condition, as in the human data.

This simulation suggested that division of labor between the two levels (implicit and explicit) of the ACS and bottom-up learning were important for explaining human performance in this task.

Synergy of Rule-Based and Similarity-Based Reasoning

The example below addresses the synergy between rule-based and similarity-based reasoning through the interaction between the top and bottom levels within the NACS.

Task and Data

In experiment 1 of Sloman (1998), participants were given pairs of arguments, each consisting of a premise statement and a conclusion statement. For example,

a. All flowers are susceptible to thrips. → All roses are susceptible to thrips.
b. All plants are susceptible to thrips. → All roses are susceptible to thrips.

Participants were asked to pick the stronger of the two arguments from each pair.

The results showed that the more similar argument from each pair of arguments was chosen much more often. Statistical tests showed that these percentages were significantly above chance, either by participants or by argument pairs (Sloman, 1998).

It should be apparent that if only rule-based reasoning (RBR; e.g., based on deductive logics) had been used, then similarity should not have made a difference because the conclusion category was contained in the premise category and thus both arguments in each pair should have been equally, perfectly strong. Therefore, the data suggested that similarity-based reasoning (SBR), as distinct from rules or logics capturing category inclusion relations, was involved to a significant extent.

In experiment 2, participants were instead asked to rate the likelihood of each argument. Ratings could range from 0 to 1. Statistical tests showed that ratings were significantly below 1, by participants and by arguments. Again, note that it would have been the case that the outcome was 1 if only RBR had been used (because the conclusion category was contained in the premise category).

In experiment 4, participants were asked to rate the likelihood of each argument. However, in this case, each category inclusion relation was specifically presented as part of each argument. The results showed that the mean judgment was almost 1. In other words, the similarity-based phenomena almost disappeared.

Experiment 5 was similar to experiment 2. However, before any rating was done, participants were asked to make category inclusion decisions. Thus, in this case, participants were reminded of rules explicitly involving category inclusion relations. Therefore, they were more likely to use RBR, although probably not as much as in experiment 4, due to the separation of category inclusion judgment and argument likelihood rating.

The results showed that none of the participants gave a likelihood judgment of 1 for every argument, indicating that SBR might be at work. Compared with experiment 2, however, having participants make category inclusion judgments earlier increased the likelihood rating, probably reflecting the increased involvement of RBR.

Simulation

In CLARION, the NACS was mainly responsible for performing reasoning tasks (through the action control by the ACS). Computationally, the following process was posited (Sun & Zhang, 2006): first, a premise statement is presented to the NACS (through the action of the ACS). Each premise statement (e.g., "all flowers are susceptible to thrips") is encoded as a rule in the top level of the NACS. The two concepts involved in it are encoded as well, as chunks, both implicitly (in distributed representations) and explicitly (as individual chunk nodes). Inclusion relations, such as "roses are flowers," already exist and are encoded also as rules at the top level. When it comes to dealing with the conclusion statement (e.g., "all roses are susceptible to thrips"), the chunk representing the first concept of the conclusion statement (p. 128) (e.g., "rose") is presented to the NACS. Due to the similarity between the first concept of the conclusion statement and the first concept of the premise statement (i.e., the similarity between the two corresponding chunks), the chunk representing the latter is partially activated. Thus, the encoded rule representing the premise statement becomes applicable. As a result of this rule application, the chunk representing the target concept (the second concept of the premise statement; e.g., "thrips") is also partially activated, the extent of which is proportional to the aforementioned similarity. When there are multiple such arguments, this process is repeated.

Experiment 1 and experiment 2 both involved SBR to a very significant extent. Experiment 4 involved explicit use of categorical relations and thus mainly RBR. Experiment 5 involved more of SBR, along with RBR. The relative emphasis of RBR versus SBR was accomplished through the balancing parameters in CLARION.

In the simulation of experiment 1, as in the human data, the more similar argument from each pair of arguments was chosen much more often. These percentages were significantly above chance, as in the human data. In this simulation, there was a significant involvement of SBR. This simulation demonstrated that the significant involvement of SBR in producing the human data of this experiment was a reasonable interpretation, given the close match with the human data.

In the simulation of experiment 2, simulated participants were to rate the likelihood of each argument. The results were significantly below 1, different from what would have been predicted if only RBR had been used, both by subjects and by arguments, the same as in the human data. This simulation again demonstrated the significant involvement of SBR, as in the human data.

In simulating experiment 4, simulated participants were to rate the likelihood of each argument with the corresponding category inclusion relation. The simulation produced a mean judgment of close to 1, the same as in the human data. Compared with experiment 2, explicit RBR based on category inclusion was much more prominent in this case, which captured the human data accurately.

In simulating experiment 5, ratings were obtained after an initial phase during which category inclusion decisions were made. In this case, participants were reminded of RBR involving category inclusion relations and therefore they were more likely to use RBR compared with experiment 2, although not as much as in experiment 4. In this simulation, the mean judgment was indeed higher than in experiment 2 but lower than in experiment 4, the same as in the human data. This simulation showed that the interpretation as embodied in the simulation setup was a reasonable one.

In all, the simulation of these experiments successfully substantiated the analysis of human performance in this task described earlier. In particular, SBR was captured through the interaction of implicit and explicit processes at the two levels of the NACS.

Human Motivation, Personality, and Emotion

Motivation

CLARION accounts for many psychological phenomena related to human motivation. For example, Lambert et al. (2003) showed that in socially stressful situations, social stereotyping was more pronounced. They used the task of recognition of tool versus gun, with priming by black or white faces. The results showed that, in socially stressful situations, when paired with a black face, tools were much more likely to be mistaken for guns. This phenomenon has been captured, explained, and simulated using CLARION. When certain avoidance-oriented drive strengths become very high within the MS, the processing within the ACS of CLARION becomes very implicit, as determined by the MCS on the basis of the drive strength levels within the MS. The implicit processing within the ACS is susceptible to stereotyping effects. The simulation using CLARION captured the corresponding human data well (Wilson, Sun, & Mathews, 2009), and provided a detailed, mechanistic, and process-based explanation for the data.

Likewise, skilled performance may deteriorate when individuals are under pressure. For example, in terms of mathematical skills, Beilock et al. (2004) showed that performance worsened when pressure was high. To demonstrate this point, they used a modular arithmetic problem set of the form *A = B (mod C)* and tested participants either under pressure (with monetary incentives, peer pressure, and social evaluation) or not. The result showed significant differences with and without pressure. This task has been simulated using CLARION, which provided detailed, mechanistic, and process-based explanations. When certain avoidance-oriented drive strengths are very high, processing within the ACS becomes very implicit (controlled by the MCS on the basis of the drive strength levels from the MS). Overly implicit processing leads to worsened (p. 129) performance (see, e.g., Sun et al., 2005). The simulation captured the corresponding human data (Wilson, Sun, & Mathews, 2009).

Personality

On the basis of the CLARION model of human motivation, human personality may be accounted for as well. The CLARION personality model is, to a significant extent, based on drives within the MS. On that basis, goal setting (by the MCS) and action selection (by the ACS) take place. Individual differences may be accounted for (for the most part) by the differences in relative drive strengths in different situations by different individuals. Individual differences in terms of relative drive strengths are consequently reflected in the resulting goals, major cognitive parameters, and action selection on that basis. Personality types, in addition to being mapped onto drive activations, may also be mapped in part onto other mechanisms and processes (although to a lesser degree). For instance, within the CLARION framework, personality may involve parameters within the ACS, NACS, MS, and MCS. Therefore, personality is the result of complex interactions among a large set of modules and processes. This approach may be justified from a variety of perspectives (see Sun & Wilson, 2014, for details).

This model of personality captures more detailed aspects of psychological processes than does previous work (e.g., Read et al., 2010; Shoda & Mischel, 1998, etc.). It goes beyond notions of goals, plans, resources, beliefs, and the like. It grounds personality traits in a cognitive architecture so that they are explained in a unified way along with many other psychological phenomena, data, and constructs, based on the primitives envisaged within the cognitive architecture.

Various simulation tests show that the CLARION personality model is capable of demonstrating stable personality traits but, at the same time, showing sufficient variability of behaviors (Sun & Wilson, 2014). It maps onto and computationally demonstrates the well-known Big Five personality structures, among other things.

This CLARION personality model has also been used to simulate and explain human data. For example, Moskowitz et al. (1994) examined the influence of social role/status on interpersonal behavior in a work environment. It was hypothesized that social role/status would have an effect on behavior. Study participants were expected to behave more submissively, for example, when interacting with a boss versus a coworker or a subordinate. They were also expected to be more dominant, for example, when with a subordinate or a coworker than with a boss. Event contingent recording was used to gather data, and the data analysis confirmed these expected effects. The CLARION simulation captured all the major effects exhibited within the human data. Various other simulations of human data have also been carried out and shed new light on psychological processes underlying personality (Sun & Wilson, 2014).

Emotion

According to CLARION, emotion is the collective result of operations throughout a system. It should not be viewed as a unitary thing. Its emergence may involve physiological states, physiological reactions, action readiness, physical (external) actions, motivational processes, evaluation/attribution processes, and meta-cognitive processes, as well as decision making and reasoning of various forms. According to CLARION, emotion is the sum total of all of these in particular circumstances.

In CLARION, emotion may be explained by a multitude of processes involving the ACS (for actions), NACS (for evaluation), MS (for motivation), and MCS (for meta-cognitive regulation). In particular, emotion is closely related to the motivational subsystem. Smillie et al. (2006), Carver and Scheier (1998), and Ortony et al. (1988) have addressed the importance of motivation and expectation in generating emotion.

For example, it has been hypothesized within CLARION that the emotion of *elation* may be related to positive reward (including "unexpected" positive reward) and also, to a lesser extent, "expectation" of positive reward. Computationally, the intensity of elation may be in part a function of drive strengths among the approach-oriented drives, which (in part) determine reward.

On the other hand, the emotion of *anxiety* may be related to "expectation" of negative reward. The intensity of anxiety may be in part a function of the strengths of avoidance-oriented drives (Smillie et al., 2006). Smillie et al. (2006) discussed the link between the avoidance system and anxiety (see also Carver & Scheier, 1998, p. 92). Furthermore, the emotion of *fear* may result from "expectation" of more intense negative reward. Computationally, the intensity of fear may be determined in part by a function of the strengths of avoidance-oriented drives, in the same way as anxiety.[2]

Emotional processes mainly occur in the bottom levels of CLARION in various subsystems (Sun & Mathews, 2012); that is, emotional processing is (p. 130) mostly implicit (although not all implicit processes are emotional; see LeDoux, 1996; Damasio, 2005). Furthermore, its locus may lie, to a large extent, in the MS, because it is closely related to motivation (as discussed earlier), and also in the ACS, because it is closely tied to action. Frijda (1986) and Arnold and Gasson (1954), for example, suggested the importance of "action readiness" in emotional experience. Explicit processes may also have some role in emotion, for example, through affecting decisions of the bottom levels or through explicit reasoning in "cognitive appraisal" (Frijda, 1986). However, they are not the main locus of emotion according to CLARION.

General Discussion

Some Possible Arguments Against CLARION

Without delving into psychological process details, some might wrongly believe that CLARION is an ad hoc collection of artificial intelligence (AI) techniques and algorithms. It should be made clear that the work is not about AI techniques or algorithms, but about cognitive/psychological processes and mechanisms. The computational techniques employed in CLARION were not randomly thrown together, but selectively included in order to account for a maximum range of psychological data and phenomena. That is, they were selected for the sake of developing a comprehensive theory of the mind. The overarching meta-principle for CLARION as a cognitive/psychological theory may be summarized as (1) minimum mechanisms, (2) effective integration, and, as a result, (3) maximum scope (i.e., maximum coverage of psychological phenomena). That is, in a sense, the development of CLARION was based on cost-benefit considerations in which the cost is the complexity of the system and the benefit is the scope of psychological data and phenomena that it is capable of capturing and explaining.

Even given this, some might argue that CLARION consists of “old” AI techniques, and, as such, there is little new. The CLARION theory, at the conceptual level, is certainly not about computational techniques. Even the CLARION computational cognitive architecture itself is, primarily, not about computational techniques (although there have been many innovations). In this regard, one should not confuse the theory (and the resulting cognitive architecture) with the tools that it employs in expressing itself. Rather, CLARION is about selectively including a minimum set of mechanisms, structured in a parsimonious but effective way, to account for a maximum set of psychological data and phenomena. The novelty of computational techniques is not a relevant issue.

Currently existing computational details in CLARION constitute an existence proof of what can be accomplished with this framework. Newer computational techniques, if proved significantly better performance-wise, can be relatively easily inserted into the architecture to replace old techniques without making major changes to the theory of CLARION.

However, some may then argue that some theoretical notions on which CLARION is based may be controversial. For instance, implicit learning is a somewhat controversial topic. In this regard, it should be pointed out that, although implicit learning research is somewhat controversial, the existence of implicit processes is generally not in question; what is in question is their extent and importance. CLARION allows for the possibility that both types of processes coexist and interact with each other, so CLARION goes beyond the controversies that focused mostly on the minute details of implicit learning.

For instance, some past criticisms of implicit learning focused on the alleged inability to isolate processes of implicit learning. Such methodological problems are not relevant, because, in CLARION, it is recognized that both implicit and explicit processes are present in most tasks and that they are likely to influence each other in a variety of ways. Another strand of past criticisms concerned the fact that implicit learning was not completely autonomous and was susceptible to the influence of explicit cues, attention, and intention. These findings are, in fact, consistent with the two interacting levels in CLARION. Generally speaking, controversies concerning such theoretical notions are not very relevant to CLARION.

Similarity and Differences with Other Models

There are many other cognitive architectures in existence. A number of major differences exist between CLARION and these cognitive architectures, due, in no small part, to the basic underlying philosophical differences and to the different eras during which they were first conceived. I discuss just a few major differences here.

One important difference is that, in these other cognitive architectures, there is no principled distinction and/or separation between implicit and explicit processes; that is, there is no principled explanation of the distinction (e.g., based on a representational difference). Ad hoc assumptions have to be made (p. 131) regarding which particular component is explicit or implicit. As a result, they do not naturally capture the psychological process of the interaction between implicit and explicit processes. They provide no direct explanation of the effects resulting from the interaction between the two as shown in empirical data (e.g., the synergy effects; see Sun et al., 2005).

Another major difference is that most of these other cognitive architectures are not meant for autonomous learning without a great deal of a priori (pre-given, hand-coded) knowledge to begin with. Similarly, they do not directly capture the psychological process of bottom-up learning due to the lack of capabilities for autonomous learning and the lack of the distinction between the two types of processes and the dual representational structure.

As a result of its dual representational structure, CLARION is capable of “automatic” and effortless similarity-based reasoning, whereas other cognitive architectures may have to use computationally costly and cumbersome pairwise similarity relations to enable SBR, which do not appear to be cognitively realistic (Sun, 1995).

In most of these other cognitive architectures, there is no sufficiently psychologically realistic, built-in modeling of motivational processes. As a result, goals are often externally set and directly hand-coded. They do not reflect the diversity, complexity, and flexibility of human motivation and behavior.

On the other hand, some of these other cognitive architectures may have some detailed sensory-motor modules that CLARION currently does not include (in the current release of the code; although such capabilities were implemented earlier in CLARION).

Note that CLARION and other cognitive architectures often account for different tasks, although there have been some overlaps (see, e.g., Sun et al., 2009).

Conclusion

This chapter covers the essentials of the CLARION cognitive architecture. CLARION is distinguished by its inclusion of multiple, interacting subsystems that are radically different from other cognitive architectures: the ACS, NACS, MS, and MCS. It is also distinguished by its focus on the separation and interaction of implicit and explicit processes (within these different subsystems). More importantly, it is distinguished by its emphasis on motivational and meta-cognitive processes and their interactions with other processes. Note that it is psychologically unrealistic to focus only on cognition (in the narrow sense) by ignoring motivational and other fundamental processes. With these mechanisms, CLARION has something unique to contribute to cognitive modeling, as shown by the examples included herein.

Future work on CLARION and hybrid cognitive architectures in general may include efforts to (1) better achieve the principle of minimum mechanisms and maximum scope through exploring a larger set of data, (2) develop more in-depth treatment of emotion and motivation, (3) deal with more varieties of meta-cognitive processes, (4) better capture sensory-motor processes, (5) address more complex real-world tasks (including robotics), and so on.

Acknowledgments

Thanks are due to Susan Chipman for her invitation to write this chapter. This work has been supported in part by the ONR grant N00014-08-1-0068 (thanks to Paul Bello). Thanks are also due to my many collaborators: Sebastien Helie, Nick Wilson, Michael Lynch, Robert Mathews, Sean Lane, and many others.

CLARION has been implemented as Java and C# libraries, available at http://www.cogsci.rpi.edu/~rsun/clarion.html. For the full technical details of CLARION, see Sun (2003), which is also available at this URL.

References

Ackerman, P., & Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition* (pp. 119-141). Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Arnold, M., & Gasson, S. (1954). Feelings and emotions as dynamic factors in personality integration. In M. Arnold & S. Gasson (Eds.), *The human person* (pp. 294–313). New York, NY: Ronald.

(p. 132) Beilock, S., Kulp, C., Holt, L., & Carr, T. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology*, *133*, 584–600.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*, 839–855.

Carver, C., & Scheier, M. (1998). *On the self-regulation of behavior*. Cambridge, England: Cambridge University Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2(10), 406–416.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 189–202.

Damasio, A. (2005). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Penguin Putnam. (Original work published 1994.)

Deci, E. (1980). Intrinsic motivation and personality. In E. Staub (Ed.), *Personality: Basic issues and current research* (pp. 35–80). Englewood Cliffs, NJ: Prentice Hall.

Domangue, T., Mathews, R., Sun, R., Roussel, L., & Guidry, C. (2004). The effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*(5), 1002–1011.

Elliot, A., & Thrash, T. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, *82*(5), 804–818.

Evans, J., & Frankish, K. (Eds.). (2009). *Two minds: Dual processes and beyond*. Oxford, England: Oxford University Press.

Frijda, N. (1986). *The emotion*. Cambridge, England: Cambridge University Press.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system* (2nd ed.). New York, NY: Oxford University Press.

Hasher, J., & Zacks, J. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*, 356–358.

Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, *117*(3), 994–1024.

Hull, C. (1951). *Essentials of behavior*. New Haven, CT: Yale University Press.

Jacoby, L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, *22*, 485–508.

Klein, S., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review*, *109*(2), 306–329.

Lambert, A., Payne, B., Jacoby, L., Shaffer, L., Chasteen, A., & Khan, S. (2003). Stereotypes and dominant responses: On the "social facilitation" of prejudice in anticipated public contexts. *Journal of Personality and Social Psychology*, *84*, 277–295.

Lane, S., Mathews, R., Sallas, B., Prattini, R., & Sun, R. (2008). Facilitative interactions of model- and experience-based processes: Implications for type and flexibility of representation. *Memory and Cognition*, *36*(1), 157–169.

LeDoux, J. (1996). *The emotional brain*. New York, NY: Simon and Schuster.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*(4), 523-530.

Maslow, A. (1943). A theory of human motivation. *Psychological Review*, *50*, 370-396.

Maslow, A. (1987). *Motivation and personality*, 3rd ed. New York, NY: Harper and Row.

McDougall, W. (1936). *An introduction to social psychology*. London, England: Methuen & Co.

McFarland, D. (1989). *Problems of animal behaviour*. Singapore: Longman Publishing.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*, 229–258.

Moscovitch, M., & Umilta, C. (1991). Conscious and unconscious aspects of memory: A neuropsychological framework of modules and central systems. In: R. Lister & H. Weingartner (Eds.), *Perspectives on cognitive neuroscience*. New York, NY: Oxford University Press.

Moskowitz, D. S., Suh, E. J., & Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *Journal of Personality and Social Psychology*, *66*, 753–761.

Montague, P. R. (1999). Review of reinforcement learning: An introduction. *Trends in Cognitive Science*, *3*(9), 360–361.

Murray, H. (1938). *Explorations in personality*. New York, NY: Oxford University Press.

Norman, K., Detre, G., & Polyn, S. (2008). Computational models of episodic memory. In: Sun, R. (Ed.), *Cambridge handbook on computational psychology,* 189-225. New York, NY: Cambridge University Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structures of emotions.* New York, NY: Cambridge University Press.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, MA: MIT Press.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General. 118*(3), 219–235.

Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review, 117*(1), 61–92.

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, *8*(3), 179–193.

Roediger, H. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*(9), 1043–1056.

Rogers, T. (2008). Computational models of semantic memory. In Sun, R. (Ed.), *Cambridge handbook on computational psychology* (pp. 226-266). New York, NY: Cambridge University Press.

Rosenbaum, D., Carlson, R., & Gilmore, R. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology*, *52*, 453–470.

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition*. Cambridge, MA: MIT Press.

Sallas, B., Mathews, R., Lane, S., & Sun, R. (2007). Developing rich and quickly accessed knowledge of an artificial grammar. *Memory and Cognition*, *35*(8), 2118–2133.

Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501–518.

(p. 133) Seger, C. (1994). Implicit learning. *Psychological Bulletin, 115*(2), 163–196.

Sloman, S. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, *35*, 1–33.

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review*, *10*, 320–335.

Squire, L. (1987). *Memory and brain*. New York, NY: Oxford University Press.

Stanley, W., Mathews, R., Buss, R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology, 41A*(3), 553–577.

Sun, R. (1995). Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence (AIJ), 75*(2), 241–296.

Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum.

Sun, R. (2003). *A tutorial on CLARION 5.0.* Technical report, RPI. Retrieved from **http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf**

Sun, R. (2009). Motivational representations within a computational cognitive architecture. *Cognitive Computation*, *1*(1), 91–103.

Sun, R. (2012). Memory systems within a cognitive architecture. *New Ideas in Psychology*, *30*, 227-240.

Sun, R., & Mathews, R. (2005). *Exploring the interaction of implicit and explicit processes to facilitate individual skill learning* (Technical Report TR-1162). Arlington, VA: Army Research Institute for the Social and Behavioral Sciences.

Sun, R., & Mathews, R. (2012). Implicit cognition, emotion, and meta-cognitive control. *Mind and Society*, *11*(1), 107–119.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*(2), 203-244.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, *112*(1), 159-192.

Sun, R., & Wilson, N. (2014). A model of personality should be a cognitive architecture itself. *Cognitive Systems Research*, 29-30, 1-30.

Sun, R., & Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, *18*(2), 169–191.

Sun, R., Zhang, X., Slusarz, P., & Mathews, R. (2007). The interaction of implicit learning, explicit hypothesis testing learning, and implicit-to-explicit knowledge extraction. *Neural Networks*, *20*(1), 34–47.

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

Tolman, E. C. (1932). *Purposive behavior in animals and men*. New York, NY: Century.

Toth, J., Reingold, E., & Jacoby, L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 290-303.

Tulving, E., & Schacter, D. (1990). Priming and human memory systems. *Science*, *247*, 301-305.

Willingham, D. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*(3), 558-584.

Wilson, N., Sun, R., & Mathews, R. (2009). A motivationally-based simulation of performance degradation under pressure. *Neural Networks*, *22*, 502–508. (p. 134)

Notes:

(1.) Note that a generalized notion of “drive” is adopted here, different from the stricter interpretations of drives (e.g., as physiological deficits that need to be reduced by corresponding behaviors; Hull, 1951; Weiner, 1982). In our sense, drives denote internally felt needs of all kinds that likely may lead to corresponding behaviors, regardless of whether the needs are physiological or not, whether the needs may be reduced by the corresponding behaviors or not, or whether the needs are for end states or for processes. Therefore, it is a generalized notion that transcends controversies surrounding the stricter notions of drive.

(2.) Generally speaking, there is no clear distinction between anxiety and fear in clinical psychology and psychophysiology research although there was some separation using pharmacological and direct lesion methods (see Smillie et al., 2006, p. 324).

Ron Sun

Ron Sun, Cognitive Sciences Department, Rensselaer Polytechnic Institute