Motivational Representations within a Computational Cognitive Architecture

Ron Sun
Rensselaer Polytechnic Institute, Troy, NY 12180, USA
rsun@rpi.edu

December 27, 2008

Abstract

This paper discusses essential motivational representations necessary for a comprehensive computational cognitive architecture. It hypothesizes the need for implicit drive representations, as well as explicit goal representations. Drive representations consist of primary drives — both low-level primary drives (concerned mostly with basic physiological needs) and high-level primary drives (concerned more with social needs), as well as derived (secondary) drives. On the basis of drives, explicit goals may be generated on the fly during an agent's interaction with various situations. These motivational representations help to make cognitive architectural models more comprehensive and provide deeper explanations of psychological processes. This work represents a step forward in making computational cognitive architectures better reflections of the human mind and all its motivational complexity and intricacy.

1 Introduction

It is not too far-fetched to posit that, to survive, a cognitive agent must meet the following criteria in its everyday activities (among others):

- Sustainability: An agent must attend to its essential needs, such as hunger and thirst, and also know to avoid physical dangers, and so on (Toates, 1986).
- Purposefulness: The action of an agent must be chosen in accordance with some criteria, instead of completely randomly (Hull, 1943; Anderson, 1993), and those criteria are related to enhancing sustainability of an agent (Toates, 1986).

- Focus: An agent must be able to focus its activities with respect to specific purposes. That is, its actions need to be somehow consistent, persistent, and contiguous, with respect to its purposes (Toates, 1987). However, an agent needs to be able to give up some of its activities, temporally or permanently, when necessary (e.g., when a much more urgent need arises; more later; Simon, 1967; Sloman, 1987).
- Adaptivity: An agent must be able to adapt its behavior (i.e., to learn) for the sake of improving its purposefulness, sustainability, and focus.

We contend that, in order to meet these criteria, motivational representations need to be formed that can address issues related to purpose and focus. Motivational dynamics is an essential part of human (or animal) behaviors. And it is ever-present in such behaviors—"Man is a perceptually wanting animal" as Maslow (1943) put it. Maslow (1943) argued that "the situation or the field in which the organism reacts must be taken into account but the field alone can rarely serve an an exclusive explanation for behavior. ...... Field theory cannot be a substitute for motivation theory."

In the remainder of this paper, first, a generic computational cognitive architecture CLARION will be briefly sketched, which provides the context in which motivational representations will be developed. Second, some essential considerations for necessary motivational representations will be discussed. Then, details of motivational representations within CLARION will be presented, including primary drives, derived drives, and explicit goals. Drives strength determinations will be discussed after that. This will be followed by a description of the overall structure of the motivational (sub)system. Some past and current simulations based on CLARION using those motivational representations will be outlined. Some final remarks conclude the paper.

2 The Context: The CLARION Cognitive Architecture

CLARION is a comprehensive framework of a variety of psychological processes, implemented computationally. It has been described in detail and justified psychologically in Sun (2002, 2003) (see also Sun et al., 2001; Sun, Slusarz, and Terry, 2005).

It is particularly worth noting that CLARION is an integrative architecture, consisting of a number of distinct subsystems (with a dual-representational structure in each subsystem: implicit versus explicit representations). Its subsystems include the action-centered subsystem (the ACS), the non-action-centered subsystem (the NACS), the motivational sub-system (the MS), and the meta-cognitive subsystem (the MCS). The role of the action-centered subsystem is to control actions, regardless of whether the actions are for external physical movements or for internal mental operations. The role of the non-action-centered subsystem is to maintain

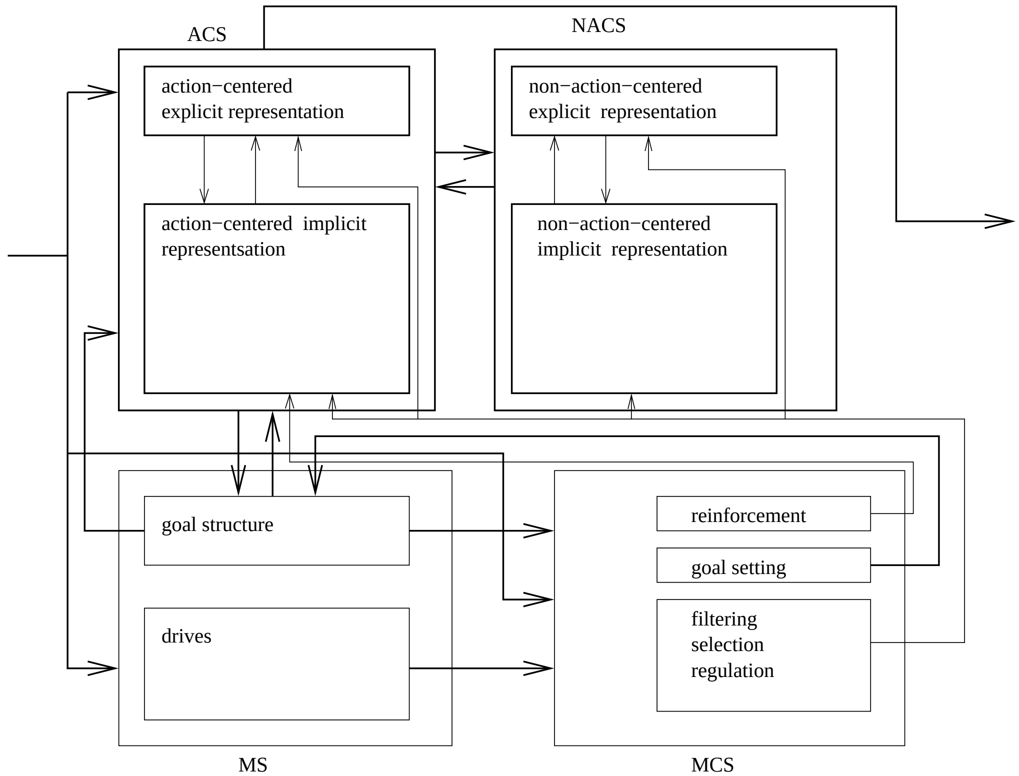

Figure 1: The CLARION Cognitive Architecture

general knowledge, either implicit or explicit. The role of the motivational subsystem is to provide underlying motivations (implicit or explicit) for perception, action, and cognition, in terms of providing impetus and feedback. The role of the meta-cognitive subsystem is to monitor, direct, and modify the operations of the action-centered subsystem dynamically (as well as the operations of the other subsystems). (See Sun, 2002; Sun et al., 2001; Sun et al., 2005; and so on for examples of the working of these subsystems.)

Each of these interacting subsystems consists of two "levels" of representation (i.e., a dual-representational structure) as posited in Sun (2002) and Sun et al. (2005): Generally, in each subsystem, the top level encodes explicit knowledge (using symbolic/localist representations, as discussed extensively in Sun, 2002) and the bottom level encodes implicit knowledge (using distributed representations, as discussed in Sun et al., 2005). The two levels interact, for example, by cooperating in action decision making, through a combination of the action recommendations from the two levels respectively, as well as by cooperating in learning through a bottom-up learning process (whereby implicit knowledge is acquired first and then explicit knowledge on its basis) and a top-down learning process (whereby explicit knowledge is acquired first and then implicit knowledge on its basis; see Sun et al., 2001; Sun et al., 2005 for psychological justifications). See Figure 1.

Some of the most important characteristics of CLARION include:

- The dichotomy of implicit and explicit cognition (which is arguably fundamental to the human mind; Reber, 1989; Sun, 2002)
- The focus on the cognition-motivation-environment interaction (Sun, 2003)
- The constant interaction of multiple components and subsystems within CLARION, with the simultaneous presence of implicit cognition, explicit cognition, motivation, meta-cognition, and so on (to capture complex, realistic psychological processes; Sun, 2003)
- Autonomous and bottom-up learning: CLARION can learn on its own, regardless of whether there is a priori or externally provided domain knowledge, while it does not preclude top-down learning or exclude innate biases, innate behavioral propensities, prior knowledge, etc. (Sun et al., 2001; Sun, 2002).

Below, we will develop necessary motivational representations within this framework, so as to be able to provide a deeper and more comprehensive explanation of behaviors.

3 Some Basic Considerations Regarding Motivational Representations

Questions concerning the mechanistic (computational) processes of motivation need to be asked. For example, how should the internal motives (e.g., needs, desires, or drives) of an agent be represented? Are they explicitly represented (as symbolic/logicist AI would suggest), or are they implicitly represented (in some way)? Are they transient, or are they relatively invariant temporally? How do contexts affect their status? How do their variations affect behaviors? A motivational (sub)system may need to be posited, which may be concerned specifically with the representations of basic needs, desires, drives, motives, and their dynamics, as well as more complex or more derivative motivational structures.

Let us examine the issue of explicit versus implicit representations of motivations within the motivational (sub)system. On the one hand, it is hard to imagine that there is no explicit representation of *goals* in a cognitive agent, since all the evidence points to the contrary (see, e.g., the theories of human skill learning by Anderson, 1993; Anderson and Lebiere, 1998; Newell, 1990). On the other hand, the internal process of drives, needs, or desires are certainly not explicit and not readily accessible cognitively (Hull, 1943; more later). So, it seems reasonable to assume that (1) the idea of dual representation (implicit and explicit) is applicable here (Sun, 2002; Sun et al., 2005) and (2), relatedly, implicit motivational processes are primary and more essential than explicit motivational processes (Sun, 2002; more discussions later).

We may further hypothesize that the explicit motivational representations consist mainly of explicit goals of an agent (Anderson and Lebiere, 1998). Explicit goals provide specific and tangible motivations for actions. Explicit goals also allow more behavioral flexibility (or "individuation" as termed by Epstein, 1982), and formation of expectancies (Epstein, 1982). While implicit motivational states may be more inclined to change from moment to moment, explicit goal representations are more persistent and longer lasting. In many circumstances, persistence is needed (as discussed before). Furthermore, it may sometimes be necessary to compute a match of a state of the world to the goal, so as to discern the progress in achieving the goal (and to generate context-dependent reinforcement signals as discussed in detail in Sun, 2003). This match may be facilitated by using an explicit representation of goals. In addition, explicit goal representations may facilitate explicit cognitive processes (in other subsystems) working on these goals and their attainment, in addition to involving implicit processes.

However, as mentioned before, the most fundamental part of the motivational (sub)system, its implicit level, consists of basic drives, basic needs, basic desires, intrinsic motives, and so on (whatever one calls them). We will refer to them all as "drives" here (Sun, 2003, 2006). In the past, Hull (1951) developed the most detailed conception of "drives"—an implicit, preconceptual representation of motives. In his view, drives arose from need states, behaviors were driven so as to eliminate need states, and drive reduction was the basis of reinforcement. Although Hull's conception of drive had significant explanatory power, his theory failed to capture many motivational phenomena—the variety of different motivations (in many organisms) proved too difficult to be encompassed by his theory of drive. A more general notion is therefore needed.

A generalized notion of "drive" is adopted here, different from the stricter interpretation of drives (e.g., as physiological deficits that require 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 (e.g., the need for a cat to engage in the process of catching mice; Herrstein, 1977). Therefore, it is a generalized notion that transcends controversies surrounding the stricter notions of drive. [1]

Turning to the dual nature of motivational representations, the motivational processes of humans are known to be highly complex and varied (see, e.g., Weiner, 1992), and ap-

[1] This notion is adopted, because we need to account for (1) context-dependent and (2) persistent but terminable drivers of behavior, (3) in an implicit way, as well as other properties of behavior mentioned early on.

parently cannot be captured with simple explicit goal representations alone (e.g., as in Anderson and Lebiere, 1998 or as in Rosenbloom et al., 1993). For example, the interactions of motives, especially their combinations, require more complex representations (McFarland, 1989; Tyrell, 1993). Their changes over time, which are often gradual and dynamic, also require a more quantitative and graded representation. Moreover, Maslow (1943) and Murray (1938) specifically discussed the unconscious characteristics of "needs". Given the above, it is natural to hypothesize that implicit motivational processes are necessary and fundamental (Heidegger, 1927; Sun, 2002, 2003). Only on that basis, explicit goal representations arise, which clarify implicit motivational (and behavioral) dynamics. Castelfranchi (2001), for example, discussed such implicit-to-explicit motivational processes, in ways analogous to general implicit-to-explicit cognitive "emergence" (as broadly discussed in Sun, 2002).

Empirical evidence from social psychology also points to the duality of human motivation. For example, Wood and Quinn (2005) explored extensively the duality of motivations in everyday life, and the relationship between implicit and explicit motivations, in ways analogous to the analysis of implicit and explicit cognitive processes in general in Sun et al. (2005). Strack and Deutsch (2005) expressed similar view, describing what I have termed top-down and bottom-up influences (implicit motivations affecting explicit motivations and vise versa; Sun et al., 2005). Aarts and Hassin (2005) reviewed evidence of both explicit and implicit motivations in human behavior. Norton et al. (2004) showed that people might be motivated implicitly by questionable criteria but then masked their implicit biases through engaging in casuistry explicitly. Woike (1995) showed how implicit and explicit motives might have different effects on memory recall. Hing et al. (2005) also demonstrated how implicit and explicit motivations might diverge and consequently how they might counter-balance each other (see also Gaertner and Dovidio, 1986). Adams et al. (1996) even found that an individual's implicit and explicit motivations could be diametrically opposed.

A bipartite motivational representation may be as follows: The (explicit) goals (such as "finding food") of an agent may be generated based on (past and current) internal drive states (for example, "being hungry") of the agent (accomplished by the meta-cognitive subsystem, to be discussed later). This explicit representation of goals derives from, and hinges upon, (implicit) drive states. [2]

Note that the dual motivational process is, to some extent, innate, having been molded by long evolutionary processes, but it allows the possibility of adjustments/adaptation from the existential experience of an individual agent (Weiner, 1992).

[2] Note that, although drive states may sometimes be identified individually (as I will do next), such identifications are approximate. They do not represent the full complexity of the matter. Furthermore, the generation and change of these drive states are fully implicitly determined (through neural networks). Thus, I view drive states as being fundamentally implicit.

4 Primary Drives in CLARION

Based on the afore-discussed considerations, a set of primary drives may be posited within CLARION (at the bottom level of the motivational subsystem; see Sun, 2003) as follows, which includes both low-level primary drives and high-level primary drives.

4.1 Low-level Primary Drives

First of all, there are the low-level primary drives (which are mostly physiological): for example, hunger, thirst, physical danger, ..., and so on. Judging from the literature on this issue, it appears justified to posit the following set of low-level primary drives (Tyrell, 1993; Lewin, 1936; Hull, 1943; McClelland, 1951; Murray, 1938):

- Food. [3]
- Water.
- Sleep.
- Avoiding physical dangers.
- Reproduction.
- Avoiding unpleasant stimuli. (Note that, although some other low-level drives may also result from unpleasant stimuli, these stimuli usually come from more specific, more identifiable sources. Hence they are separately listed as individual low-level primary drives. See Murray (1938).)

There are also drives for other physiological needs, such as physical exercise (Reiss, 2004), avoiding boredom, and so on.

It should be noted that many other physiological needs may be reduced to some of these drives above or their combinations (e.g., avoiding overly high or low temperature, urination, defecation, and so on may be attributed to the drive to avoid unpleasant or repulsive stimuli; see Murray, 1938).

As shown before (e.g., by Neubery et al., 2005; Reiss, 2004; Clancey et al., 2006), these presumably evolutionarily hard-wired low-level primary drives influence human behavior in everyday life in a significant way.

[3] This drive may be further differentiated as there may be different needs for different nutrients in accordance with bodily states; “tastes” are changeable over time.

4.2 High-level Primary Drives

Beyond low-level drives (concerning mostly physiological needs), there are also higher-level drives (which are mostly social). Some of these high-level drives are primary, in the sense of being innate or “hard-wired”. High-level primary drives may include motivations for seeking of social approval, striving for social status, desire for reciprocation, interest in exploration, ... and so on.

Judging from the existing literature (see Murray, 1938; Reiss, 2004; Maslow, 1987; James, 1890; McDougall, 1936), it appears safe to posit the following set of high-level primary drives within CLARION:

- Affiliation and belongingness. According to Murray (1938), it denotes the need to “form friendships and associations. To greet, join, and live with others. To co-operate and converse sociably with others. .... To join groups”. It is essentially the same as the need for social contact proposed by Reiss (2004). It is also similar to the notion of belongingness as proposed by Maslow (1987). As Maslow put it, it denotes “our deep animal tendencies to herd, to flock, to join, to belong.” Clearly, this drive is species-specific—not all species have an equally strong need for social belongingness.

- Recognition and achievement. It is the need to “excite praise and commendation. To demand respect. To boast and exhibit one’s accomplishments. To seek distinction, social prestige, honours or high office”. And to “overcome obstacles, ... to strive to do something difficult as well and as quickly as possible”. Maslow claimed that “all people in our society .... have a need or desire for a stable, firmly based, usually high evaluation of themselves, for self respect or self esteem, and for the esteem of others”. It includes the desire for competence, adequacy, and so on. Murray (1938) referred to them as the need for superiority.

- Dominance and power. This is similar to the notion of power proposed by Reiss (2004). and the notion of dominance proposed by Murray (1938). According to Murray (1938), it denotes the need to “influence or control others. To persuade, prohibit, dictate. To lead and direct. To restrain. To organize the behaviour of a group”.

- Autonomy. According to Murray (1938), it is the need to “resist influence or coercion. To defy an authority or seek freedom in a new place. To strive for independence”. See also Reiss (2004). Like some other drives, this drive is species-specific—not all species have an equally strong need for autonomy.

- Deference. “To admire and willingly follow a superior.... To co-operate with a leader. To serve gladly” (Murray, 1938).

- Similance. “To empathize. To imitate or emulate. To identify oneself with others. To agree and believe” (Murray, 1938).

- Fairness. Evolutionary psychology (e.g., Barkow et al., 1992) has suggested that people have a fairness instinct that prompts one to seek fairness in social interactions (including in economic activities). It is certainly related to the notion of vengeance of Reiss (2004), which is the desire to get even. It appears that the notion of vengeance may be derived from the drive for fairness (as well as possibly other drives).

- Honor. Similar to the notion of honor proposed by Reiss (2004), it denotes the desire to obey a moral or cultural code. See also the need for blame-avoidance in Murray (1938).

- Nurturance. It is the need to “mother” a child and the need to help the helpless (Murray, 1938). See also the need for family proposed by Reiss (2004).

- Conservation. “To arrange, organize, put away objects. To be tidy and clean”. And to “collect, repair, clean and preserve things” (Murray, 1938). See also the notion of order and the notion of saving in Reiss (2004).

- Curiosity. It is the desire for knowledge (Reiss, 2004), and the need to “explore.... To ask questions. To satisfy curiosity. To look, listen, inspect” (Murray, 1938).

Note that the notion of “drive” here refers to the desire to act in accordance with some *perceived* deficits or needs, which may or may not be physiological, and the act may or may not lead to the reduction of the perceived deficits/needs (cf. Hull, 1951). Thus, it is a generalized notion that provides essential underlying motivations for action (in a fundamentally implicit and embodied fashion).

In empirical research, it has been shown that these drives identified above are largely uncorrelated with each other, with only a few exceptions (as summarized in Reiss, 2004). Thus, it is reasonable to view them as (relatively) independent drives.

For each individual, each of these drives may be “weighted” somewhat differently when deciding on goals and actions, thus leading to individual differences (Reiss, 2004; Sun and Wilson, 2009). The difference in relative drive strengths or “weights” is an important source of personality differences (although not necessarily the only source; Sun and Wilson, 2009).

For each of these drives, there is often a desirable level of satisfaction that is neither the highest nor the lowest; that is, there is a “moderate mean” (as termed by Aristotle, 1953; Reiss, 2004) that is being sought after (which, nevertheless, may vary from individual to individual). For example, one may often seek a moderate amount of food, a moderate degree of power, and so on. Discrepancy between that “moderate mean” and the currently obtained amount may (partially) determine the strength of the corresponding drive for an individual.

Over-saturation may lead to unpleasant or repulsive stimuli (see the drive to avoid unpleasant or repulsive stimuli mentioned earlier).

Note that the list of motivations here may not be complete. However, this framework is meant to illustrate how a set of well worked-out motivational constructs can explain a complex and psychologically important range of behaviors (e.g., see the various psychological simulations discussed later).

4.3 Justifications of High-level Primary Drives

Comparing this set of hypothesized drives with Murray's proposal (1938), one can see that they are essentially the same, with only a few differences. For example, the drive for conservation in our framework covers both the need for conservance and the need for order proposed by Murray. The need for retention in Murray's framework may be derived from the drive for conservation in our framework. Murray's acquisition need may also be derived from the need for conservation. Murray's need for inviolacy may be attributed to the drive for recognition and achievement, as well as the drive for dominance and power, according to our framework.

Some other needs identified by Murray, such as contrarience, aggression, abasement, rejection, succorance, exposition, construction, and play, are not fundamental (primary) needs or drives in our view—they are likely the results of more fundamental (i.e., primary) drives or their combinations. For example, the need for play may be attributed sometimes to the drive of curiosity, and sometimes to the physiological drive of avoiding boredom or avoiding repulsive or unpleasant stimuli (e.g., when over-work leads to work-related stimuli becoming unpleasant). For another example, Murray's contrarience need, if exists, may be attributed to the drive for recognition and achievement and/or the drive for dominance and power. [4]

Comparing this set of hypothesized drives with Reiss' proposal (2004), we note that they are highly similar, but with some noticeable differences. For example, the need for saving and the need for order as proposed by Reiss (2004) are included in the drive for conservation in our framework. The need for family as proposed by Reiss, in our view, may be derived from the drive for affiliation and belongingness, as well as the drive for nurturance and the drive for honor. In Reiss (2004), vengeance includes the desire to get even, which, in our view, is derived from the drive for fairness and the drive for honor in our framework; vengeance in Reiss (2004) also included desires to compete and win, which may be derived from the drive for recognition and achievement, the drive for honor, and so on, We do not include Reiss' "idealism", because it may be derived from other drives in our framework (such as affiliation and belongingness, honor, fairness, nurturance, etc.). The need for status proposed by Reiss

[4] Murray's low-level (physiological, or viscerogenic in Murray's term) needs are not included in this list either. They may be attributed to some combinations of the low-level primary drives as enumerated earlier.

may be derived from the drive for dominance and power and the drive for recognition and achievement in our framework. The need for acceptance in Reiss may be derived from the drive for affiliation and belongingness, the drive for honor, and the drive for recognition and achievement in our framework. [5]

Comparing this set of hypothesized drives with McDougall (1936), we note that McDougall's framework was concerned with "instincts", not basic needs (primary drives). Instincts refer to (more or less) evolutionarily hard-wired (i.e., innate) behavior patterns/routines that can be relatively easily triggered by pertinent stimuli in pertinent situations, while basic needs (primary drives) are essential driving forces of behaviors. Instincts are different from basic needs (primary drives), because one does not have to follow instincts when there is no pertinent stimulus, and even when pertinent stimuli are present, one may be able to refrain from following them (at least more easily than from basic needs or primary drives). In other words, they are not needs, but pre-set routines — while they are relatively easily triggered, they are not inevitable. For example, William James (1890) listed the following instincts: imitation, emulation or rivalry, pugnacity/anger/resentment, sympathy, hunting, fear, appropriation/acquisitiveness, constructiveness, play, curiosity, sociability and shyness, secretiveness, cleanliness, modesty and shame, love, jealousy, parental love, ..., and so on. See also a similar list by McDougall (1936). As evident from the list above, many of these instincts are (i.e., result directly from) drives in our framework (such as "curiosity" and "parental love"), or are derived, by some means, from the drives in our framework (such as "play" and "constructiveness"). Some other instincts are not because they do not represent basic needs (e.g.,"hunting" or "jealousy").

Schwartz's (1994) 10 universal values, although addressing a different aspect of human behaviors (i.e., human "values"), bear some resemblance to the primary drives identified here. Moreover, each of these values can be derived from some primary drive in our framework or some combination of these primary drives.

Note that a number of criteria were hypothesized by Reiss (2004) regarding what constituted a drive (or a basic desire as he called it). They included: (1) each is genetically different with a different evolutionary history; (2) satiation of each produces an intrinsically valued feeling of joy; (3) each produces a different joy; (4) each is applicable to animals as well as humans (with some exceptions). These criteria are somewhat hard to verify.

Nevertheless, there have been some empirical efforts at validating some of these drives. Reiss (2004) summarized large-scale studies that had people (from different walks of life) rate the importance of each of the more than 300 motivational terms and then analyzed the factors

[5] We do not include here the need for eating, the need for tranquility, the need for physical exercises, and the need for romance, as in Reiss (2004), since they are mostly physiological (see the list of low-level primary drives earlier).

within. These studies led to a set of 16 factors, which was highly similar to the set of drives presented earlier.

Moreover, Reiss (2004) showed that results from some instruments for measuring some of these drives (as presented earlier) correlated well with other measures intended to gauge the same (or similar) constructs. For example, Reiss Profile power and order scales correlated .55 and .60 with the dominance and order scales of the PRF. Also, religiosity was found to be associated with high Reiss Profile scores for honor and family but with low scores for vengeance and independence (autonomy). Athleticism was found to be associated with traits for social contact, family, vengeance, power, but low curiosity. Reiss (2004) found that his 16 basic desires (similar to our set of drives) were largely uncorrelated to each other, which might lend support for this type of framework in general. In addition, such a framework of drives (i.e., intrinsic motivations, or basic desires) has been applied in psychopathology and mental retardation research. [6]

5 Derived Drives in CLARION

While primary drives are (more or less) hard-wired (i.e., innate) and relatively unalterable, there may also be "derived" drives. They are secondary, more changeable, and acquired mostly in the process of satisfying primary drives. Derived drives may include: (1) gradually acquired drives, through "conditioning" (Hull, 1951); (2) externally set drives, through externally given instructions. For example, due to the transfer of the desire to please superiors into a specific desire to conform to his/her instructions, following a certain instruction may become a (derived) drive.

6 Explicit Goals in CLARION

On top of implicit and embodied motivations, namely drives, explicit goals may also be present, which are more specific and more explicit (when compared with drives). Explicit goals may be set (by the meta-cognitive subsystem) based on (primary or derived) drives (Simon, 1967; Nerb et al., 1997). The drives provide the context within which explicit goals are set and carried out.

For details regarding explicit goal representations, see Sun (2003; in particular, the chapter on the action-centered subsystem). Briefly, a goal list consists of a number of goal slots, each

[6] The empirical evidence appears to support multiple similar frameworks, not necessarily any particular one. The differences among these frameworks may be adjudicated through empirical and theoretical means, in particular, through capturing, simulating, and explaining empirical data using computational models (Sun, 2009).

of which can hold a goal (along with its parameters). These goals on the list compete to be the current goal (based on activation levels). The current goal is then used in action decision making (by the action-centered subsystem).

Explicit goals are different from drives in many respects. Specifically, (1) there may be multiple drives being activated at the same time (e.g., being hungry and being thirsty at the same time). However, there is usually only one goal being pursued at a time (Anderson and Lebiere, 1998), although a goal may encode multiple action objectives, that is, having multiple (parameter) dimensions (see Sun, 2003 for details). (2) Drives are more diffused in focus, while goals are often more specific (McFarland, 1989; Anderson and Lebiere, 1998). (3) Drives are more implicit, while goals are explicit (Murray, 1938; Maslow, 1943; Hull, 1951). (4) Drives are often hardwired, while goals are more flexibly created, set, and carried out (Hull, 1951; Sun, 2003).

7 Drive Strengths and Their Combinations

A set of essential considerations concerning drive strengths have been identified (Tyrell, 1993; Sun, 2003):

- *Proportional activation.* The activation (i.e., strength) of a drive should be proportional to the corresponding perceived deficit in a relevant aspect (such as food or water).
- *Opportunism.* Considerations concerning opportunities need to be incorporated when calculating desirability of alternatives. For example, the availability of water may lead to preferring drinking water over gathering food, provided that the food deficit is not too much greater than the water deficit.
- *Contiguity of actions.* There should be a tendency to continue the current action sequence, rather than switching to a different sequence, to avoid the overhead of switching. In particular, actions to satisfy a drive should persist beyond minimum satisfaction, that is, beyond a level of satisfaction barely enough to reduce the strength of the most urgent drive to be slightly below those of the other drives. For example, one should not run to a water source and drink only a minimum amount, and then run to a food source and eat a minimum amount, then going back to the water source to repeat the cycle.
- *Interruption when necessary.* However, when a much more urgent drive arises (such as "avoiding physical dangers"), actions for a lower-priority drive (such as "sleep") may be interrupted.
- *Combination of preferences.* The preferences for a certain course of action (i.e., a certain action goal) resulting from different drives should be combined to generate a somewhat

higher overall preference. Moreover, a compromise candidate may be generated that is not the best for any single drive but the best in terms of the combined preferences.

Let us see how these considerations may be fulfilled. First of all, the first two considerations together point to the use of products, such as *food-deficit * food-stimulus*, in determining strengths of drives, which takes into consideration both deficit and availability (Tyrell, 1993).

The next consideration necessitates a persistent goal structure, as discussed earlier, which can be set and then persist until an interruption by a much more urgent drive (such as "avoiding physical dangers" when a danger is close by). In this way, we may avoid "thrashing": switching back and forth among two or more alternative tasks that are demanded by drives with comparable strengths, while preserving the possibility of interruption when a much more urgent need arises.

Combination of preferences when deciding on a goal is an issue that deserves careful consideration. It is believed that combination should be carried out by the resemblance of a multi-vote voting system whereby a goal emerges from tallying the multiple votes cast by different drives (cf. Tyrell, 1993). The problem with the single-vote approach is that only the top-priority goal of each drive is taken into consideration, but lesser goals may be ignored, which may nevertheless make excellent compromise candidates. The multi-vote approach takes into consideration multiple preferences. Following this approach, we may implement the combination of preferences in a variety of ways. For example, a connectionist network may be used to implement a multi-vote approach, which leads to the setting of a goal based on all the preferences of all the active drives.

Finally, in relation to what has been discussed above, a few examples of calculating drive strengths are as follows:

- *Food*. As mentioned before, the strength of this drive is determined by two factors: *food-deficit* felt by the agent, and the *food-stimulus* perceived by it. The involvement of "stimulus" in determining the strength is necessary, because otherwise an agent may be dominated by one slightly larger deficit and ignore availability issues all together. [7] Thus, the product of *food-deficit * food-stimulus* is included. However, *food-deficit* alone needs to be taken into account too, because otherwise an agent may starve to death if *food-stimulus* is not available at all (while, e.g., *water-stimulus* is abundantly available). Note that *food-stimulus* captures both the "incentive" value of a food item as well as its accessibility (its distance) and certainty (see Hull, 1951). Thus, the strength of this drive may be set to *0.95 ** max *(0.30 * food-deficit, food-deficit * food-stimulus)*, where $0 \leq$ *food-deficit* ≤ 1 and $0 \leq$ *food-stimulus* ≤ 1. The maximum strength of this drive,

[7] For example, when water is nearby and easily accessible, and *food-deficit* is not too much greater than *water-deficit* but food stimulus is not available, the agent should address the *water-deficit* first.

in this case, is thus 0.95 (although parameters may vary from individual to individual).

- *Water*. This drive is similar to the drive for *food*. For the same reason as described above, both *water-deficit* and *water-deficit * water-stimulus* should be taken into account in the determination of the strength. Thus, the strength of this drive may be set to *0.95 ** max *(0.30 * water-deficit, water-deficit * water-stimulus)* (although parameters may vary from individual to individual).

- *Avoiding physical dangers*. The strength of this drive is proportional to the danger signal: its distance, severity (disincentive value), and certainty. These factors may be captured by *danger-stimulus* (which is presumably determined by distance, severity, and certainty). Thus, the strength of this drive may be set to *0.98 * danger-stimulus * safety-deficit*, where "safety-deficit" measures the internal sensitivity of an individual with regard to danger (where $0 \leq$ *safety-deficit* ≤ 1 and $0 \leq$ *danger-stimulus* ≤ 1). The maximum strength of this drive, in this case, is thus 0.98 (although parameters may vary from individual to individual).

- *Sleep*. The strength of this drive may be determined based on physical exhaustion as well as by night proximity (or similar stimuli). Thus, the strength of this drive may be something along the line of *0.95 * sleep-stimulus * sleep-deficit*.

- *Reproduction*. This drive is always present to a certain extent (determined by internally felt "deficit"), and intensifies when *mate-stimulus* is present, proportional to the intensity of *mate-stimulus*, modulated by the internally felt "deficit". The strength of this drive thus may be determined by *0.90 * max (0.30 * mate-deficit, mate-stimulus * mate-deficit)* (where parameters may vary from individual to individual).

A few examples of the strengths of high-level primary drives are as follows.

- *Affiliation and belongingness* denotes "our deep animal tendencies to herd, to flock, to join, to belong" (Maslow, 1987). Its strength may be determined based on *0.70 * max (0.30 * AB-deficit, AB-deficit * AB-stimulus)* (although parameters may vary due to individual differences). The maximum strength of this drive, in this case, is thus 0.70. The strength is determined by two factors: the pertinent internally felt "deficit" (*AB-deficit*, measuring the internal inclination of an individual for affiliation and belongingness), modulated by pertinent stimuli and an ever-present component (0.30).

- *Recognition and achievement* includes the desire for competence, adequacy, recognition, attention, and so on. Its strength may be determined based on *0.50 * max (0.20 * RA-deficit, RA-deficit * RA-stimulus)* (although parameters may vary due to individual differences). The maximum strength of this drive, in this case, is thus 0.50. The

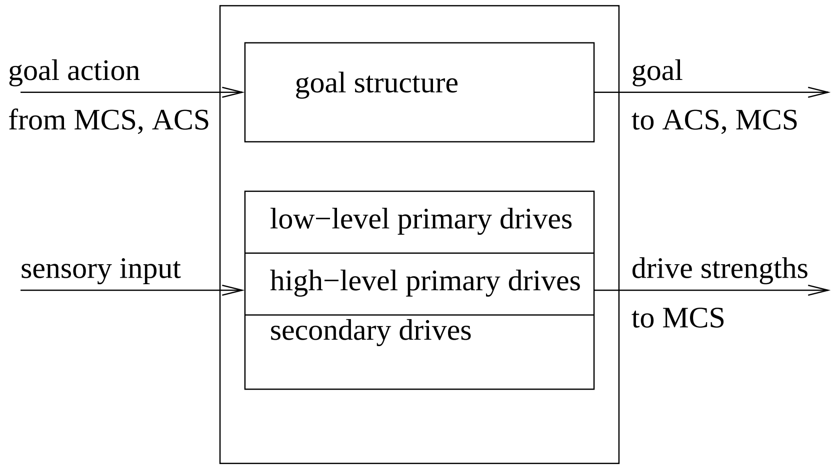

Figure 2: Structure of the motivational subsystem.

strength is determined by two factors: the pertinent internally felt "deficit" (*RA-deficit*), modulated by pertinent stimuli and an ever-present component (0.20).

Strengths of other drives may be similarly determined based on a similar set of factors.

8 Structure of the Motivational Subsystem in CLARION

The structure of the motivational subsystem (the MS) is shown in Figure 2. Note that this subsystem is not standalone—it is closed tied to the meta-cognitive subsystem (e.g., for the sake of goal setting by the meta-cognitive subsystem) and the action-centered subsystem (to set, to change, and to carry out goals by its actions).

In this subsystem, the goal structure has been described as belonging to the action-centered subsystem (Sun, 2003) —in fact, it is an integral part of both subsystems, as well as closely tied to the meta-cognitive subsystem. So it is at the center of the whole architecture. In this subsystem, the goal structure constitutes an explicit representation of motivations, and drives an implicit one. However, it is not necessarily the case that the two types of representations directly correspond to each other (e.g., one being extracted from the other; Sun, 2002, 2003).

The mapping between the state of the world (stimuli as perceived by a cognitive agent, and the sensing of various perceived deficits) and the strengths of various drives can be implemented, in accordance with the afore-specified value ranges and relations, by backpropagation networks. The networks identify relevant features (and their intensities), such as *food-stimulus*, *water-deficit*, or *mate-stimulus*, from raw sensory input. The output of such a network may be the strengths of drives. Such a network captures the implicitness of drive generation (due to the distributed representations used in the hidden layers of backpropagation networks).

In advance of cognitive modeling of specific tasks, a drive network may be pre-trained (off-line) as follows: The input to the net consist of raw sensory input, that is, without pre-processing that identifies various deficits and stimuli (although such pre-processing is possible, can be incorporated, and may make learning easier). The output are proper drive strengths, for example, as calculated from the afore-given formulas (although this is not necessarily the case). Through backpropagation learning, the network *learns* to identify relevant deficits and stimuli through its three-layered weight structure and to output proper drive strengths. It may be argued that this pre-training is a (very rough) approximation of a long evolutionary process that has gradually shaped a drive system. [8]

Furthermore, reinforcement learning is possible for "online" tuning of the drive strengths. Reinforcement signals received while an agent is interacting with the world may be used not only for learning actions to be performed by the agent (given the current state and the current goal, within the action-centered subsystem), but also for tuning drive strengths given that particular context (within the motivational subsystem), as well as for adjusting goal setting given the context (within the meta-cognitive subsystem). Relevant parameters may be adjusted in accomplishing this. Further details may be found in Sun and Wilson (2009).

9 Some Simulations Involving Motivational Representations and Processes

Many psychological simulations using CLARION involving the afore-discussed motivational representations have been undertaken, are currently under way, or have been planned. Let us look into some of them very briefly (for details, refer to the relevant citations below).

First of all, Lambert et al. (2003) showed that in socially stressful situations, social stereotyping was more pronounced. To demonstrate this point, they examined the task of the recognition of tool versus gun, with priming by black or white faces. The results show that, in socially stressful situations, when paired with a black face, tools are much more likely to be mistaken as guns. This phenomenon has been captured, explained, and simulated using the motivational representations within CLARION along with the other mechanisms of CLARION. When certain (avoidance-oriented) drive strengths become extremely high, the processing within the action-centered subsystem of CLARION becomes very implicit (which is controlled and determined by the meta-cognitive subsystem of CLARION on the basis of drive strength levels within the motivational subsystem). The implicit processing is more

[8] Note that this is the preferred alternative to using the afore-specified formulas directly, which would require various deficits and stimuli to be identified individually.

likely to be subject to stereotyping effects. The simulation using CLARION captured the corresponding human data well (see Wilson, Sun, and Mathews, 2009), including, in particular, the stereotyping effect under pressure, and provided a detailed, mechanistic, and process-based explanation for the phenomenon.

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 (using monetary incentives, peer pressure, and social evaluation) or not. The result showed clear differences between with and without pressure. This task has been simulated using the motivational subsystem, the meta-cognitive subsystem, and the action-centered subsystem within CLARION, which provide detailed, mechanistic, and process-based explanations. When certain (avoidance-oriented) drive strengths are very high, the processing within the action-centered subsystem becomes very implicit (controlled by the meta-cognitive subsystem on the basis of drive strength levels). Overly implicit processing leads to worsened performance (see, e.g., Sun et al., 2005). The simulation using CLARION captured the corresponding human data (see Wilson, Sun, and Mathews, 2009).

The same phenomenon has been captured by CLARION simulations in low-level skill domains (involving mostly sensorimotor skills). Beilock and Carr (2001) showed that golf putting performance was worse when participants were under pressure (due to video taping and other setups that induced pressure). This phenomenon has been simulated using the same mechanisms and processes as the simulations of the corresponding phenomena in high-level skill domains (as discussed above). The resulting simulation successfully captured the human data (Wilson et al., 2009).

Along a separate line, it was found that the goal of getting at the truth and the goal of getting along with others led to different styles of cognitive processing (Chen et al., 1996): With the goal of getting at the truth, systematic processing often takes place, while with the goal of getting along, shallower processing often happens (leading to agreeing with others' opinions). This phenomenon may be captured and explained using CLARION with its motivational representations. In CLARION, different drives and consequently different goals lead to different behaviors: With high drives for understanding and getting at the truth ("curiosity" as well as "recognition and achievement"), systematic processing takes place (as a result of the corresponding goal); in contrast, with high drives for getting along with others ("similance"), shallower processing happens, leading to agreeing with others' opinions.

It was also found that participants who believed that their tasks were important exhibited a great deal of systematic processing (regardless whether the messages they were supposed to process were congruent or not), while other participants showed systematic processing only when the messages they received were incongruent (Maheswaran and Chaiken, 1991). This phenomenon is again captured and explained using CLARION (with its motivational representations). Participants who believed that their tasks were important were more likely to have higher drive strengths related to getting things done right ("conservation" and "recognition and achievement"), and therefore they exhibited more effortful systematic processing (that is, regardless whether the messages they were supposed to process were congruent or not). In contrast, participants who did not believe that their tasks were important would have lower relevant drive strengths, and therefore they spent less effort and showed less systematic processing; only when the received messages were incongruent, higher drive strengths would be generated for resolving the incongruence, which then led to more systematic processing for these participants.

Yet another interesting finding is that the effects of extrinsic rewards differ for participants with high Need for Cognition (NFC) scores versus those with low NFC scores in terms of effort — time spent on a brainstorming task (Thompson et al., 1993). Such effects may be simulated and explained using CLARION with its motivational representations. The difference between these two groups of participants may be captured by motivational (drive) differences between them. From that basis, behavioral differences result. For those with high drives for intellectual endeavors (e.g., "curiosity"), extrinsic rewards distracted from that motivation (with motivations for material gains), and led to less focus on intellectual goals (but more on goals for material gains) and consequently led to spending less time on the task. In contrast, for those with low drives for intellectual endeavors, extrinsic rewards led to drives for material gains and consequently to a higher likelihood of setting the goal to work on the task (for the sake of material gains), and thereby led to spending more time on the task on average.

It was found that people sometimes made judgment based on questionable criteria, but then engaged in casuistry to mask biased decision making (Norton et al., 2004). Such findings may be captured and explained using CLARION. According to CLARION, during initial decision making, people set up goals to make decisions that promote self-interest in some way, due to many drive states that inevitably emphasize self-interest of some kind (such as "dominance and power", "recognition and achievement", "conservation", etc.). However, when they need to justify their decisions, their goals become those that emphasize socially acceptable rationales, due to some other relevant drives being highly activated (such as "honor" — obeying social norms). Thus, they try to recruit apparently justifiable (but often fictitious) reasons to support their decision making. CLARION can simulate such cases, using its motivational structures as well as other mechanisms.

Maner et al. (2005) shows that different motivations can bias a person's perception (in particular, of functionally relevant emotional expression in goal-relevant social targets). For example, activating a self-protection motive led to perceiving more anger in the faces of certain racial minority groups. Activating a mate-search motive led to perception of more sexual arousal in opposite-sex targets. This finding can be readily simulated and explained within CLARION. In CLARION, a motivational state (including drives and goals in the motivational subsystem) leads to filtering input information (e.g., for the action-centered subsystem) in accordance with the motivational state, which is accomplished by the meta-cognitive subsystem. So, the finding is consistent with CLARION.

Iyengar and Lepper (1999) shows that whether freedom of making choices leads to higher satisfaction and better performance is culture-specific. In their experiments, Anglo-American children showed less intrinsic motivation when choices were made for them, while Asian-American children demonstrated higher motivation under the same circumstances (whether choices were made for them by authority figures or by peers). This finding can be explained and simulated within the framework of CLARION. According to CLARION, the former group has higher drives for "autonomy", while the latter group has higher drives for "deference" and "similance" (all within the motivational subsystem). Effort levels and therefore performance are (in part) determined by the meta-cognitive subsystem in accordance with the motivational states (from the motivational subsystem): When the situation is congruent with the current motivational state, more effort is exerted (e.g., in the form of the meta-cognitive subsystem dictating a higher proportion of effortful explicit processing in the action-centered subsystem, which often leads to better performance; Sun et al., 2005); on the other hand, when the situation is incongruent with the motivational state, less effort is given to the task. Therefore, naturally, the former group performed better when given more choices, while the latter performed better when given less choices.

All of the afore-discussed motivational representations have been used in developing a comprehensive personality model within CLARION. The personality model essentially maps different human personality types onto different parameter values for determining drive strengths, thereby grounding personality theories in the motivational subsystem of CLARION, along with some necessary functions of the meta-cognitive subsystem (for setting goals and regulating behaviors based on drive strengths and so on) and the action-centered subsystem. Various simulations and tests show that this CLARION personality model is capable of demonstrating stable personality traits but at the same time showing sufficient variability of behaviors in response to different situations (Sun and Wilson, 2009). It maps onto, and computationally demonstrates, the well known Big Five personality theory (Read and Miller, 2002), among other things.

10 Concluding Remarks

This paper has addressed the essential motivational representations necessary for a comprehensive cognitive architecture. Within this discussion, the need for implicit drive representations, as well as explicit goal representations, has been hypothesized. Drive representations consist of primary drives (both low-level and high-level primary drives), as well as derived (secondary) drives. On the basis of drives, explicit goals may be generated on the fly during an agent's interaction with various situations.

The afore-discussed motivational representations help to make a computational cognitive architecture more complete and functioning in a more psychologically realistic way. We believe that this work constitutes a requisite step forward in making computational cognitive architectures more realistic models of the human mind taking into considerations all of its complexity and intricacy, especially in terms of its motivational dynamics.

Acknowledgments

This work has been supported in part by the ARI contract W74V8H-05-K-0002 (to Ron Sun and Bob Mathews) and the ONR grant N00014-08-1-0068 (to Ron Sun). Nick Wilson conducted some of the simulations briefly mentioned here.

References

H. Aarts and R. Hassin, (2005). Automatic goal inference and contagion. In: J. Forgas, K. Williams, and S. Laham (eds.), (2005). *Social Motivation: Conscious and Unconscious Processes*. Cambridge University Press, New York.

H. Adams, L. Wright, and B. Lohr, (1996). Is homophobia associated with homosexsual arousal? *Journal of Abnormal Psychology*, 105 (3), 440-445.

J. R. Anderson, (1993). *Rules of the Mind*. Lawrence Erlbaum Associates. Hillsdale, NJ.

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

Aristotle, (1953). *The Nicomachean Ethics*. Penguin, New York. (Originally published in 330 BC.)

J. Barkow, L. Cosmides and J. Tooby, (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press. New York.

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

S. Beilock and T. Carr, (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology*, 130, 701-725

C. Castelfranchi, (2001). The theory of social functions: Challenges for computational social science and multi-agent learning. *Cognitive Systems Research*, special issue on the multidisciplinary studies of multi-agent learning (ed. Ron Sun), 2(1), 5-38.

S. Chaiken and Y. Trope (eds.), (1999). *Dual Process Theories in Social Psychology.* Guilford Press, New York.

S. Chen, D. Shechter, and S. Chaiken, (1996). Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology*, 71 (2), 262-275.

W. Clancey , M. Sierhuis, B. Damer, and B. Brodsky, (2006). Cognitive modeling of social behaviors. In: R. Sun (ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation.* Cambridge University Press, New York.

A. Epstein, (1982). Instinct and motivation as explanations for complex behavior. In: D. W. Pfaff (ed.), *The Physiological Mechanisms of Motivation.* Springer-Verlag, Berlin.

S. Gaertner and J. Dovidio, (1986). The aversive form of racism. In: J. Dovidio and S. Gaertner (eds.), *Prejudice, Discrimination, and Racism*, pp.61-86. Academic Press, Orlando, FL.

M. Heidegger, (1927). *Being and Time.* English translation published by Harper and Row, New York. 1962.

R. Herrstein, (1977). The evolution of behaviorism. *American Psychologist*, 32, 593-603.

L. Hing, G. Chung-Yan, R. Grunfeld, L. Robichaud, and M. Zanna, (2005). Exploring the discrepancy between implicit and explicit prejudice. In: J. Forgas, K. Williams, and S. Laham (eds.), (2005). *Social Motivation: Conscious and Unconscious Processes.* Cambridge University Press, New York.

C. Hull, (1943). *Principles of Behavior: An Introduction to Behavior Theory.* D. Appleton-Century Company.

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

S. S. Iyengar and Lepper, M. R. (1999). Rethinking the role of choice: A cultural perspective on intrinsic motivation. Journal of Personality and Social Psychology, 76, 349-366.

W. James, (1890). *The Principles of Psychology.* Dover, New York.

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

J.K. Maner, Kenrick, D. T., Neuberg, S. L., Becker, D. V., Robertson, T., Hofer, B., Delton, A., Butner, J., and Schaller, M. (2005). Functional projection: How fundamental social motives can bias interpersonal perception. *Journal of Personality and Social Psychology*, 88, 63-78.

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

A. Maslow, (1987). *Motivation and Personality.* 3rd Edition. Harper and Row, New York.

D. Maheswaran and S. Chaiken, (1991). Promoting systematic processing in low-motivation settings: Effects of incongruent information on processing and judgment. *Journal of Personality and Social Psychology*, 61 (1), 13-25.

D. McClelland, (1951). *Personality.* New York: Dryden Press.

W. McDougall, (1936). *An Introduction to Social Psychology.* Methuen & Co., London.

D. McFarland, (1989). *Problems of Animal Behaviour.* Longman Publishing, Singapore.

H. Murray, (1938). *Explorations in Personality.* Oxford University Press, New York.

J. Nerb, H. Spada and A. Ernst, (1997). A cognitive model of agents in a common dilemma. *Proceedings of the 19th Cognitive Science Conference*, 560-565. Erlbaum, Mahwah, NJ.

S. Neuberg, D. Kenrick, J. Manor, and M. Shaller, (2005). From evolved motives to everyday mentation. In: J. Forgas, K. Williams, and S. Laham (eds.), (2005). *Social Motivation: Conscious and Unconscious Processes.* Cambridge University Press, New York.

A. Newell, (1990). *Unified Theories of Cognition.* Harvard University Press, Cambridge, MA.

M. Norton, J. Vandello, and J. Darley, (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87 (6), 817-831.

S. J. Read and Miller, L. C. (2002). Virtual personalities: A neural network model of personality. *Personality and Social Psychology Review*, 6, 357-369.

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

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

S. Schwartz, (1994). Are there universal aspects of human values? *Journal of Social Issues*, 50, 19-45.

A. Sloman, (1987). Motives, mechanisms and emotions. *Emotion and Cognition*, 1, 217-234

F. Strack and R. Deutsch, (2005). Reflection and impulse as determinants of conscious and unconscious motivation. In: J. Forgas, K. Williams, and S. Laham (eds.), (2005). *Social Motivation: Conscious and Unconscious Processes.* Cambridge University Press, New York.

R. Sun, (2002). *Duality of the Mind.* Lawrence Erlbaum Associates, Mahwah, NJ.

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

R. Sun (ed.), (2008). *The Cambridge Handbook of Computational Psychology.* Cambridge University Press, New York. 2008.

R. Sun, (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research*, in press.

R. Sun, E. Merrill, and T. Peterson, (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science.* Vol.25, No.2, 203-244.

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

R. Sun and N. Wilson, (2009). A computational personality model within a comprehensive cognitive architecture. Submitted.

E. Thompson, S. Chaiken, and J. Hazlewood, (1993). Need for cognition and desire for control as moderators of extrinsic reward effects: A person x situation approach to the study of intrinsic motivation. *Journal of Personality and Social Psychology*, 64 (6), 987-999.

F. Toates, (1986). *Motivational Systems.* Cambridge University Press, Cambridge, UK.

B. Weiner, (1992). *Human Motivation: Metaphors, Theories, and Research.* Sage, Newbury Park, CA.

N. Wilson, R. Sun, and R. Mathews (2009). Performance under pressure: A cognitive architecture perspective. Submitted.

B. Woike, (1995). Most memorable experiences: Evidence for a link between implicit and explicit motives and social cognitive processes in everyday life. *Journal of Personality and Social Psychology*, 68, 1081-1091.

W. Wood and J. Quinn, (2005). Habits and the structure of motivation in everyday life. In: J. Forgas, K. Williams, and S. Laham (eds.), (2005). *Social Motivation: Conscious and Unconscious Processes.* Cambridge University Press, New York.