Chapter 1

The Motivational Subsystem

Ron Sun

1.1 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 serious 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 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 (Simon 1967, Sloman 1986).
- 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. As indicated by Maslow (1943), "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."

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 ways)? Are they transient, or are they relatively invariant temporally? How do contexts affect their status? How do their variations affect performance? A motivational (sub)system may need to be posited, which may be concerned specifically with the representations of the basic needs, desires, drives, motives, and their dynamics, as well as the more complex or more derivative motivations.

Let us examine the issue of explicit versus implicit representations of motivational constructs 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 learning by Anderson 1993, Anderson and Lebiere 1998, Rosenbloom et al 1993). On the other hand, the internal process of generating drives, needs, or desires are certainly not explicit and not readily accessible cognitively (Hull 1943). So, it seems reasonable to assume that (1) the idea of dual representation is applicable here (Sun 2002) and (2), relatedly, implicit motivational processes are primary and more essential than explicit processes (Sun 2002; see more discussions later).

We may further hypothesize that the explicit motivational representation consists 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 drive states may change from moment to moment, explicit goal representations are more persistent and longer lasting. In many circumstances, persistence in goal attainment is needed (as will be discussed later in relation to Tyrell's (1993) considerations). 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 will be discussed with regard to the MCS). 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) to work on these goals and their attainment, in addition to using 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, pre-conceptual 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; Herrnstein 1977). Therefore, it is a generalized notion that transcends controversies surrounding the stricter notions of drive. 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.

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 apparently 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 drives, especially their combinations (McFarland 1989, Tyrell 1993), require more complex representations. 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 pointed out and discussed the unconscious characteristics of "needs". Given the above, it is natural to hypothesize that implicit motivational processes are necessary and fundamental. Only on that basis, the 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 discussed in Sun 2002).

Empirical evidence from social psychology also pointed 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 we 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 (which is tied to the working of the ACS, as explained before) may be generated based on (past and current) internal drive states (for example, "being hungry") of the agent (accomplished within the MCS, to be explained later). This explicit representation of goals derives from, and hinges upon, (implicit) drive states. [1]

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.

1.2 Primary Drives

Based on the afore-discussed considerations, a set of primary drives may be posited as follows, which includes both low-level primary drives and high-level primary drives.

[1] Note that, although drive states may be identified individually (as we will do next), such identifications are approximate. They do not represent the full complexity of the matter. Thus, we view drive states as being fundamentally implicit.

1.2.1 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. (This drive may be further differentiated as there may be different needs for different nutrients in accordance with bodily states; "tastes" are also changeable over time.)
- Water.
- Sleep.
- Avoiding physical dangers.
- Reproduction.
- Avoiding unpleasant stimuli. (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), these presumably evolutionarily hard-wired low-level primary drives influence human behavior in everyday life in a significant way.

1.2.2 High-level Primary Drives

Beyond low-level drives (concerning mostly physiological needs), there are also higher-level drives. Some of these high-level drives are primary, in the sense of being innate or "hard-wired". High-level primary drives (which are mostly social) may include: 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), it appears safe to posit the following set of high-level primary drives:

- 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 interaction (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 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, we believe that it is reasonable to view them as independent drives.

For each individual, each of these drives may be "weighted" somewhat differently, thus leading to individual differences (Reiss 2004).

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, etc. Discrepancy between that "moderate mean" and the currently obtained amount may (partially) determine the level 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 coverage of motivations here may not be complete. However, this framework illustrates how a set of well thought-out motivational constructs can contribute to a rather complex and interesting range of behaviors (see the simulations).

1.2.3 Justifications of High-level Primary Drives

Comparing this set of hypothesized drives with Murray's (1938), we note 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, in our framework.

Some other needs identified by Murray, such as contrarience, aggression, abasement, rejection, succorance, exposition, construction, play, are not fundamental needs or drives in our view—they are likely the results of various combinations of more fundamental (i.e., primary) drives. 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.

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.

Comparing this set of hypothesized drives with Reiss (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 our framework. 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 (such as affiliation and belongingness, honor, fairness, nurturance, etc.) in our framework. The need for status proposed by Reiss 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. 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).

Comparing this set of hypothesized drives with McDougall (1936), we note that McDougall's framework was concerned with "instincts",

not basic needs (i.e., 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 (or primary drives), because one does not have to follow instincts when there is no pertinent stimulus, and 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. So, 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 in fact 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 constitutes a drive (or a basic desire as he called it). They include: (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 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 may lend support for this type of framework in general. In addition, such frameworks of drives (intrinsic motivations, or basic desires) have been applied in psychopathology and mental retardation research.

1.3 Derived Drives

While primary drives are more or less hard-wired (innate) and relatively unalterable, there are also "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 the instructions becomes a (derived) drive. [2]

1.4 Explicit Goals

On top of implicit and embodied motivations, namely drives, explicit goals may also be used, which are more unique, more specific, and more explicit (when compared with drives). Explicit goals may be set 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 goal representations, see Sun (2003; in particular, the chapter on the ACS).

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 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 more explicit

[2] Although we do not describe derived drives in detail, one may, however, use such drives in simulation. One may (1) set up additional drives (manually) to simulate the *effect* of derived drives, or (2) add the process of acquiring derived drives to the motivational subsystem, analogous to some of the learning processes in the ACS and NACS.

(Murray 1938, Maslow 1943, Hull 1951). (4) Drives are often hardwired, while goals are more flexibly formed and carried out (Hull 1951, Sun 2003).

1.5 Considerations Concerning Drive Strengths and Their Combinations

Tyrell (1993) identified a set of considerations concerning drive strengths. These considerations include:

- *Proportional activation.* The activation of a drive should be proportional to the corresponding perceived offset (deficit) in a related aspect (such as food or water).
- *Opportunism.* Considerations concerning opportunities need to be incorporated when calculating desirability of alternatives in choosing actions. 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.
- *Persistence.* Similarly, 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 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 preference.

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 two considerations necessitate 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 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 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.

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 *food-stimulus* in determining the strength is necessary, because otherwise an agent may be dominated by one slightly larger deficit and ignore accessibility issues all together. [3] Thus, the product of *food-deficit * food-stimulus* should be 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). [4] 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 is thus 0.95.
- *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)*.
- *Avoiding physical dangers*. The strength of this drive is proportional to the danger signal: its distance, severity (disincentive value), and certainty. The first two factors may be captured by *danger-stimulus*

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

[4] Note that *food-stimulus* captures both the "incentive" value of a food item as well as its accessibility (its distance). See Hull (1951).

(which is presumably determined by distance and severity), and the third factor by *danger-certainty*. Thus, the strength of this drive may be set to *0.98 * danger-stimulus * danger-certainty*. The maximum strength of this drive is thus 0.98.

- *Sleep*. The strength of this drive may be determined based on physical exhaustion as well as by night proximity. Thus, the strength of this drive may be something along the line of *0.95 ** max *(night-proximity, exhaustion)*.
- *Reproduction*. This drive is always present to a certain extent, and intensified when *mate-stimulus* is present, proportional to the intensity of *mate-stimulus*. The strength of this drive thus may be determined by *0.30 + (0.60 * mate-stimulus)*.

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 *0.20 + (0.50 * AB-deficit)* (although it may vary due to individual differences). The maximum strength of this drive is thus 0.70. The strength is determined by two factors: the pertinent deficit felt (*AB-deficit*), and an ever-present component (0.20).
- *Recognition and achievement* includes the desire for competence, adequacy, recognition, attention, and so on. Its strength may be determined based on *0.10 + (0.40 * RA-deficit)* (although it may vary due to individual differences). The maximum strength of this drive is thus 0.50. The strength is determined by two factors: the pertinent deficit felt (*RA-deficit*), and an ever-present component (0.10).

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

1.6 Structure of the Motivational Subsystem

The structure of the motivational subsystem (the MS) is shown in Figure 1.1. [5]

In this subsystem, the goal structure has been described as belonging to the ACS (Sun 2003) —in fact, it is an integral part of both subsystems, as well as closely tied to the MCS. So it is at the center of the whole system. In this subsystem, the goal structure constitutes an explicit representation of motivations, and drives an implicit one. However, it is not necessarily

[5] Note that this subsystem is not standalone—it is closed tied to the MCS (e.g., for the sake of goal setting by the MCS) and the ACS (to set, to change, and to carry out goals).

the case that the two types of representations directly correspond to each other (e.g., one being extracted from the other), as in the case of the ACS or the NACS (Sun 2002, 2003).

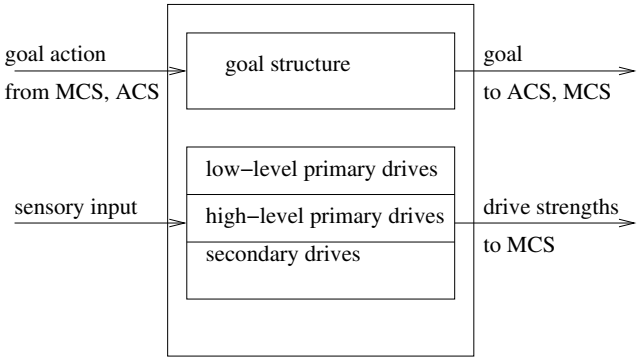

Figure 1.1. Structure of the motivational subsystem.

The mapping between the state of the world (as perceived by a cognitive agent, including 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 can be used to 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 directly the strengths of drives. Such networks capture the implicitness of drive generation (due to the distributed representations used in the hidden layers).

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 the 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 (pre-determined) drive strengths. It may be hypothesized that this pre-training is a (very rough) approximation of a long evolutionary process that has gradually shaped a drive system. (Note that this is the preferred alternative to using the afore-specified formulas directly, which would require various deficits to be given as input or identified individually.)

1.7 References

Aristotle, (1953). *The Nicomachean Ethics*. Penguin, New York. (Orginally 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.

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

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

A. Maslow, (1962). *Toward a Psychology of Being*. Van Nostrand, Princeton, NJ.

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

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

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.

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

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