7 Affect and Proto-Affect in Effective Functioning

ANDREW ORTONY, DONALD A. NORMAN, AND WILLIAM REVELLE

We propose a functional model of effective functioning that depends on the interplay of four relatively independent domains, namely, affect (value), motivation (action tendencies), cognition (meaning), and behavior (the organism's actions). These domains of functioning all need to be considered at each of three levels of information processing: the reactive, the routine, and the reflective levels. The reactive level is primarily a hard-wired releaser of fixed action patterns and an interrupt generator, limited to such things as processing simple stimuli and initiating approach and avoidance behaviors. This level has only proto-affect. The routine level is the locus of unconscious, uninterpreted expectations and well-learned automatized activity, and is characterized by awareness, but not self-awareness. This level is the locus of primitive and unconscious emotions. The reflective level is the home of higher-order cognitive functions, including metacognition, consciousness, and self-reflection, and features full-fledged emotions. In this framework, we characterize personality as a self-tunable system comprised of the temporal patterning of affect, motivation, cognition, and behavior. Personality traits are a reflection of the various parameter settings that govern the functioning of these different domains at all three processing levels. Our model constitutes a good way of thinking about the design of emotions in computational artifacts of arbitrary complexity that must

perform unanticipated tasks in unpredictable environments. It stresses the need for these artifacts, if they are to function effectively, to be endowed with curiosity and expectations and to have the ability to reflect on their own actions.

What does it take for an organism to function effectively in the world? What would a comprehensive model of the fit between an organism's functioning and the environmental conditions in which the organism finds itself look like? What role does affect play in effective functioning? Our general answer to these questions is that, for organisms of any complexity, effective functioning depends on the interplay of four domains: *affect,* what the organism feels; *motivation,* what the organism needs and wants; *cognition,* what it knows, thinks, and believes; and *behavior,* what it does.

For us, *behavior* refers only to physical action,[1] both externally observable (e.g., movements of the limbs or facial muscles) and internal (e.g., contractions of the gut or changes in heart rate). Just as the cognitive areas of the cortex are largely separable from the motor areas, we believe that, from a functional perspective, cognitive activity such as thinking and problem solving needs to be treated separately from motor activity. Cognitive activity, or *cognition,* is essentially concerned with meaning. This is in contrast to affect, which has to do with value (positive or negative). We use the term *affect* as a superordinate concept that subsumes particular valenced conditions such as emotions, moods, feelings, and preferences. Emotions are that subset of affective conditions that are *about* something, rather than being vague and amorphous, as are, for example, moods (Clore & Ortony, 2000). We also distinguish emotions from feelings. We take feelings to be readouts of the brain's registration of bodily conditions and changes—muscle tension, autonomic system activity, internal musculature (e.g., the gut), as well as altered states of awareness and attentiveness. Emotions are *interpreted* feelings, which means that feelings are necessary but not sufficient for emotions. This definition is different from that of Damasio (2000), who views emotion itself as the registration of the bodily changes and the feeling (of an emotion) as a mental image of those changes. We prefer our view to Damasio's because we think that emotions proper have cognitive content, whereas feelings themselves do not; thus, we view feelings as components of emotions rather than the other way around. The last domain of functioning, *motivation,* concerns tendencies to behave in certain ways—in particular, to attain or avoid certain kinds of state, such as satiation, danger, or becoming successful.

A central organizing theme of our discussion is that affect and the other domains of functioning need to be considered at each of three levels of information processing: the *reactive*, the *routine*, and the *reflective* (Fig. 7.1). One of our main claims is that affect manifests itself in different ways at the different levels of processing. We believe that viewing affect and its relation to information processing in this way helps to resolve some of the debates about affect and emotion.

Some of the more important differences between the three levels are presented in Table 7.1. The main function of the most elementary level, the reactive level, is to control the organism's approach and avoidance behavior and, as described by Sloman, Chrisley, and Scheutz (see Chapter 8), to interrupt and signal higher levels. At this level, there is only simple, unelaborated affect, which we refer to as "proto-affect." The realm of proto-affect is restricted to the here and now, as opposed to the future or the past.

The second, routine, level is primarily concerned with the execution of well-learned behaviors. At this level, affect begins to show some of the features of what we would ordinarily call emotions but in a rather limited and primitive manner. These "primitive emotions" can involve information relating to the future as well as to the present. For example, simple forms of hope and fear necessitate some minimal form of expectation. We consider

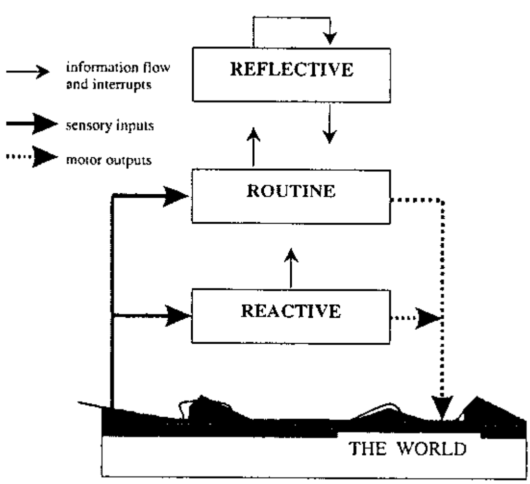

Figure 7.1. Simplified schematic of the three processing levels, *reactive*, *routine*, and *reflective*, showing their principal interconnections and relationships to one another and to the state of the world. Small solid lines represent information flow and interrupt signals that serve to indicate activity; broken lines indicate response initiation; thick lines indicate sensory signals.

Table 7.1. Principal Organism Functions at Three Levels of Information Processing

	Processing Level		
	Reactive	Routine	Reflective
Perceptual input	Yes	Yes	No
Motor system output	Yes	Yes	No
Learning	Habituation, some classical conditioning	Operant and some classical conditioning, case-based reasoning	Conceptualization, analogical, metaphorical, and counterfactual reasoning
Temporal representation	The present and primitive representation of the past	The past, present, and primitive representation of the future	The past, present, future, and hypothetical situations

the states discussed by Fellous and LeDoux (see Chapter 4) or the states related to reinforcement discussed by Rolls (see Chapter 5) to be routine-level, "primitive" emotions.

Finally, the third and most sophisticated level, reflective, is the locus of higher-level cognitive processes and consciousness. At this level, we get full-fledged emotions that are cognitively elaborated; that can implicate representations of the present, the future, or the past; and that can be named. These are the kinds of emotional state that are the focus of appraisal theories (e.g., Arnold, 1960; Lazarus, 1966; Mandler, 1984; Ortony, Clore, & Collins, 1988; Roseman, 1984; Scherer, 1984).

We are certainly not the first to propose a multilevel analysis of information processing. Many others have proposed such accounts, although often starting from quite different places. For example, Broadbent (1971), considering evidence of similarities and differences in the effects of various stress manipulations, argued for at least two levels of cognitive control; and Sanders (1986) discussed how multiple levels of energetic and cognitive control (including arousal, activation, and effort) are utilized as a function of task demands (see also Revelle, 1993). Advocates of computational approaches have also proposed models involving several levels of information processing (e.g., Sloman & Logan, 2000; also Sloman et al., Chapter 8, and Minsky, in preparation), representing impressive and highly elaborated examples. Our approach is in the same spirit as these, focusing as it does on the implications of the information-processing levels we identify for a model of affect and effective functioning.

The model that we propose is a functional one. However, we believe that many aspects of it are consistent with neuroanatomical accounts, with the three levels—the reactive, the routine, and the reflective—corresponding roughly to the assumed functions of the spine/midbrain basal ganglia, cortex, cerebellum, and prefrontal cortex. It thus bears some similarity to MacLean's (1990) early proposal of the triune brain, which although in many ways problematic is still a useful framework (see Fellous and LeDoux, Chapter 4). Proposals such as those of Fellous (1999) and of Lane (2000) are also compatible with our general approach.

As already indicated, we consider affect to be a general construct that encompasses a wide range of psychological conditions relating to value. However, even though emotions are more highly specified than other affective states, they do not comprise a discrete category with easily identifiable boundaries. Rather, they vary in their typicality, with some cases being better examples than others. Thus, we propose that the best examples of emotions, which we often refer to as "full-fledged emotions," are interpretations of lower-level feelings and occur only at the reflective level, influenced by a combination of contributions from the behavioral, motivational, and cognitive domains. At the middle, routine, level, we propose basic feelings, "primitive emotions," which have minimal cognitive content, while at the most elementary, reactive, level of processing, we argue that there are no emotions at all. All that is possible at the reactive level is an assignment of value to stimuli, which we call "proto-affect." This in turn can be interpreted in a wide range of ways at higher levels from a vague feeling that something is right or wrong (routine level) to a specific, cognitively elaborated, full-fledged emotion (reflective level). Although these different kinds of affective state vary in the degree to which they involve components of prototypical emotions, many are still emotion-like, albeit not very good examples (Ortony, Clore, & Foss, 1987). The graded nature of the concept of emotion is readily accommodated by our account partly because the specificity of an affective state is held to depend on the information-processing level at which it appears.

Although our focus in this chapter is on affect and emotion, we generally consider the affective aspects of an organism's functioning in light of their interactions with the other domains of functioning (behavior, motivation, and cognition) because we believe that this sort of integration is important. We start with a discussion of these issues in the context of biological beings and then consider how some of these ideas might apply to the design of robots and autonomous agents. For convenience, we sometimes discuss the levels within an organism (e.g., a human or a robot) and sometimes in terms of different organisms (e.g., organisms that might be restricted in the number of levels available to them). We discuss our analysis as it relates to personality and individual differences. Since we take affect, motivation, and cognition to be the internal control mechanisms of behavior, we view differences in their steady-state parameters as comprising the strikingly consistent and strikingly different organizations that are known as personality. So, for example, an individual who exhibits strong positive affect and strong approach behaviors might be classified by a personality theorist as an *extravert*. Such a person responds very differently to environmental inputs than one—an *introvert*—who has a strong negative affect system and strong inhibitory behaviors.

Our goal here is to lay out a general framework that might help us in thinking about a number of issues relating to affect and emotion. Thus, when we discuss some particular function or behavior in the context of one or other of the levels of processing, we are not necessarily rigidly committed to the idea that the function is performed at, or exclusively at, that level. In a number of instances, our assignments of functions to levels are speculative and provisional. For example, classical conditioning encompasses a wide range of behaviors and learning which may not all involve the same brain structures (e.g., some involve the hippocampus and some do not). Accordingly, some of the phenomena of classical conditioning should probably be thought of as originating from only the reactive level and some from the routine level, but we are undecided about exactly how to conceptualize the distribution of these phenomena across levels.[2] Readers should understand, however, that we are more concerned with articulating a way of thinking about how affect, motivation, cognition, and behavior interact to give rise to effective functioning than we are with particular details. It is our hope that future research will enable some of our proposals to be tested empirically. In the meantime, we are finding them to be a fruitful way of reconceptualizing some issues related to effective functioning in general, as well as issues in more specific domains such as personality theory, people's responses to products, and the design of autonomous, intelligent systems.

AFFECT AT THREE LEVELS OF PROCESSING

In this section, we discuss the way in which affect is manifested at each level of processing. The reactive level is primarily a releaser of fixed action patterns and an interrupt generator (Fig. 7.1). These interrupts are generally registered at the next level up, the routine level, which is the locus of well-learned automatized activity, characterized by awareness but not self-awareness. Self-awareness arises only at the highest, reflective, level, which is the home of higher-order cognitive functions, including metacognition, consciousness, and self-reflection. We interpret the existing psychological and neurological evidence to indicate that the reflective level processes are prefrontal, which means that, unlike the reactive and routine levels, the reflective level neither receives direct sensory inputs nor directly controls motor output. Reflection is limited to analyzing internal operations and to biasing and otherwise controlling routine-level activity. In fact, both reactive- and routine-level processing can modulate the operating characteristics of the reflective level, for example, by changing attentional focus, both by patterns of neural firing and chemically, through neurotransmitter changes.

Affect at the Reactive Level: Proto-Affect

Reactive-level processing comprises biologically determined responses to survival-relevant stimuli and is thus rapid and relatively unsophisticated with respect to both its detection mechanisms and its behavioral repertoire. New activity at the reactive level generally results in modification of output, but none of the activity is cognitive in nature—there is no cognition at the reactive level. Furthermore, at the reactive level, the other three domains of functioning—affect, behavior, and motivation—are so closely intertwined that they are better thought of as different perspectives on the same phenomenon rather than different phenomena. Reactive-level behavioral responses are of two broad classes—approach and avoidance—each governed by mechanisms of activation and inhibition. These responses also serve as alerting mechanisms, interrupting and causing higher levels of processing (in organisms that have them) to attend to the interrupting event and thus sometimes permitting a better course of action than would otherwise have been possible. The sophistication of the reactive level varies with the sophistication of the organism so that amoebas, newts, dogs, and humans vary considerably in the range of stimuli to which their reactive levels are responsive as well as in the types of behavior that reactive-level processes can initiate. Most reactive-level processing is accomplished through pattern recognition, a mechanism which is fast but simple and thus limited in scope. This means that it has a high potential for error, in both false diagnoses (false alarms) and missed ones (misses). The associated behaviors—motor responses—are either very simple, such as reflexes or simple fleeing or freezing behaviors, or preparatory to more complex behaviors governed by higher information-processing levels. In rare cases, reactive-level behaviors can involve more coordinated responses such as those necessary for maintaining balance.

The reactive level is highly constrained, registering environmental conditions only in terms of immediately perceptible components. Consequently, it needs and has a very restricted, simple, representational system; in particular, its crude and limited representations of the past are restricted to those that are necessary for habituation and simple forms of classical conditioning. In particular, registration of anomalous events is highly restricted, limited to such things as local violations of temporal sequencing. Nevertheless, although processing that necessitates comparing a current event with past events (e.g., case-based reasoning) is unavailable to reactive-level processes, in complex organisms such as humans, a great deal can still be accomplished through the hard-wired, reactive-level mechanisms.

So far, our discussion of reactive-level processes has focused on behavioral responses. This is because there is much less to say about reactive-level motivation and affect. The only forms of motivation that are operative at the reactive level are simple drives (e.g., appetitive and survival drives). Given a modicum of evolutionary complexity, organisms can have multiple drives that are sometimes incompatible. For example, the newt, motivated to copulate (below the water), is also motivated to breathe (above the water); sometimes the behavior can be modified to accommodate both drives, but otherwise, the more critical will dominate and temporarily inhibit the other (Halliday, 1980).

As for affect, our proposal is that at the reactive level there is only the simplest form of affect imaginable, what we call "proto-affect." For all of the stimuli that the organism encounters, the reactive level assigns values along two output dimensions, one of which we call "positive" and the other "negative."[3] These signals, which are the fundamental bases of affect and emotion, are interpreted by and interrupt activity at higher levels of processing. Thus, proto-affect represents nothing more than the assignment of valence to stimuli. At the same time, these reactive-level affective signals are so intimately related to behavioral (especially motor) responses and to the motivation to approach or avoid a stimulus that it makes little sense to try to distinguish them from one another. Throughout evolutionary history, the specificity of the automatic response systems has grown so much that, in the human, there are specific, prepared responses to a wide range of stimulus classes. For example, smiling faces, warm environments, rhythmic beats, and sweet tastes automatically give rise to predominantly positive valence, while frowning faces, extreme heat or cold, loud or dissonant sounds, bitter tastes, heights, and looming objects immediately induce negative valence.

Reactive-level processes enervate the motor system in preparation for one of a limited set of fixed action pattern responses. Consider, for example, how a human responds to the taste of a bitter or caustic substance. On our analysis, the reactive level assigns negative value to the substance, so the motivation is to immediately reject it. The body withdraws, the mouth puckers, the diaphragm forces air through the mouth, ejecting the food. At the same time, human observers typically attribute specific affective states based on their observations of such behavior. For example, if, as observers, we see a baby grimace, move its head away, and spit out a substance, we say that the baby "dislikes" the substance. However, in our analysis, at the reactive level, all that exists is proto-affect and the tightly coupled motivation to expel the substance and the associated behavior of spitting it out. Because as observers we see this nexus of motivation and behavior, we attribute an emotional state to the baby—we talk about the baby disliking the substance. However, our view is that an emotion of dislike or disgust involves an interpretation that can take place only at higher levels of processing.[4]

Of course, reactive-level responses to any given stimulus are not identical across different people (or even in the same person on all occasions). In other words, the parameters governing the operating characteristics of reactive-level functioning can vary across individuals and, although generally to a lesser degree, within individuals from one time to another. The kinds of parameter that we have in mind here include the strength, speed, accuracy, and sensitivity of a variety of basic functions carried out at the reactive level. For example, it is likely that the strengths of approach and avoidance behaviors vary and interact with activation and inhibition, which also vary in strength. Variations in sensitivities might be expected of, for example, perceptual acuity and anomaly detection (as in the detection of temporal irregularities or discontinuities). In addition, we assume that the latency and intensity of signals sent to higher levels of processing vary.

Affect at the Routine Level: Primitive Emotions

The core of the routine level is the execution of well-learned routines—"automatic" as opposed to "controlled" processes (e.g., Schneider & Shiffrin, 1977). In contrast to the reactive level, the routine level is capable of a wide range of processes, from conditioning involving expectancies to quite sophisticated symbolic processing. This is the level at which much human behavior and cognition is initiated and controlled. Here, elementary units are organized into the more complex patterns that we call "skills." Behavior at the routine level can be initiated in various ways, including activity at the reflective level (e.g., deciding to do or not to do something) and, as just discussed, activity at the reactive level (Fig. 7.1). Some routine-level processes are triggered by other routine-level activity. Finally, some routines can be triggered by the output of sensory systems that monitor both internal and external signals.

As well as routinized behavior, the routine level is the home of well-learned, automatic cognitive processes, such as the cognitive aspects of perception and categorization, basic processes of language comprehension and production, and so on. Indeed, we call this level "routine" because it encompasses all non-reactive-level processing that is executed automatically, without conscious control (which is the purview of the reflective level). However, although there is no consciousness at the routine level, awareness is an important cognitive aspect of it. Earlier, we defined cognition as the domain associated with meaning and affect as the domain associated with value. One of the things of which we can be (cognitively) aware is (affective) feeling; but although there is awareness at the routine level, there is no self-awareness. This is because self-awareness is a reflexive function. Since routine-level processes cannot examine their own operations, self-awareness is possible only at the reflective level.

Expectations play an important role at the routine level. Routine-level processes are able to correct for simple deviations from expectations, although when the discrepancy becomes too large, reflective-level control is required. Consider the case of driving an automobile. If the routine level registers a discrepancy between the implicit expectations and what actually happens and if the driver has sufficient expertise, the routine level can quickly launch potential repair procedures, even though such procedures might sometimes be suboptimal. On the other hand, an inexpert driver may have no routine procedures at all to engage under such conditions, in which case slower reflective-level processes take control and generate a conscious decision as to what to do, which might well be too late. In other words, there is a speed/accuracy tradeoff at work with respect to the two levels.

It is important to emphasize that routine-level expectations are implicit rather than explicit. They are the automatic result of the accumulation of experiences that forms a general model of likely, or "normal," outcomes or events—stored norms which are automatically recruited when anomalies occur (Kahneman & Miller, 1986). The strength of these expectations together with the intensity of valence associated with the current and expected states influence the strength of the ensuing feelings. Expectations might also arise from some kind of continuity-of-experience mechanism—an implicit belief that the future is not apt to deviate much from the recent past. However, whether the expectations are learned or rooted in expectations of experiential continuity, the key point is that the routine level can only detect expectation violations: it cannot interpret them. Only the reflective level can interpret and understand discrepancies and their consequences and then provide active, conscious decisions as to what to do about them. When some discontinuity, potential problem, or disruption of a normal routine is encountered, an interrupt is generated that, in its turn, generally launches other processes and affective reactions. This interrupt might be thought of as a primitive form of surprise. However, in the model, this interrupt is not valenced; therefore, it is categorically not an affective signal or emotion of any kind (Ortony & Turner, 1990). The system needs to do more work before value can be assigned. Thus, we view surprise as the precursor to emotion (Mandler, 1984). This is consistent with the neuroanatomical finding (Kim et al., 2004) that while one region of the amygdaloid complex responds similarly to fear and surprise (suggesting that valence has not yet been assigned), a separate region is responsive to fear but not surprise (suggesting that it is responding to valence per se).

Whereas the reactive level can have only unelaborated positive and negative affect, some minimal elaboration does occur at the routine level. Given our view that the routine level allows for some representation of the future as well as the present, four elementary cognitive categories emerge as a result of crossing these two levels of time with the two levels of valence (positive and negative). These four categories lie at the heart of the rudimentary, primitive emotions that arise at the routine level. In terms of the kinds of emotion specification described by Ortony, Clore, and Collins (1988), these four experientially discriminable primitive emotional states can be characterized as follows:

1. A (positive) feeling about a **good thing** (present)
2. A (negative) feeling about a **bad thing** (present)

3. A (positive) feeling about a **potential good thing** (possible future)
4. A (negative) feeling about a **potential bad thing** (possible future)

If we were to try to assign conventional emotion names to these states (which we think is inadvisable), the first two could be said to correspond roughly to something like "happiness" and "distress" and the second two to primitive forms of "excitement" and "fear," respectively.[5] We call these "primitive emotions," to convey the idea that they are routine-level feelings—affective states which have not yet been interpreted and cognitively elaborated. We think that animal studies of the kind reported by LeDoux (1996) and studies with humans involving unconscious processing of fear-relevant stimuli (e.g., Öhman, Flykt, & Lundqvist, 2000) are studies of routine-level, primitive emotions. As we discuss in the next section, there is an important difference between the "primitive" fear of the routine level and fully elaborated fear, which occurs only at the reflective level. Our analysis, in which four of the primitive emotions result from the product of two levels of valence (positive and negative) and two levels of time (present and future), is also consistent with the proposals of researchers such as Gray (1990), and Rolls (1999; see Chapter 5).

We propose that affective states at the routine level have some, but not all, of the features of a full-fledged emotion and that, at this level, affective states are related to but separable from cognition and motivation. The routine level lacks the cognitive resources necessary to interpret feelings as emotions by making the kind of rich, conscious elaborations of situations (e.g., reasoned, causal attributions) that characterize full-fledged emotions. Sophisticated processes such as these are available only at the reflective level.

We now need to consider the nature of motivation at the routine level. Whereas at the reactive level we had only simple motivations such as drives and approach-and-avoidance tendencies, much richer motivational structures, such as inclinations, urges, restraints, and other, more complex action tendencies, guide behavior at the routine level. These motivations to engage in or inhibit action are now clearly distinct from the actions themselves and related to, but again clearly distinct from, primitive emotions. At the reactive level, motives are entirely driven by cues, whether internal or external, but the motivation disappears when the cue goes away.[6] In contrast, at the routine level, motivations persist in the absence of the associated cue, dissipating only when satisfied. A good historical example of this is the Zeigarnik effect (Zeigarnik, 1927/1967), wherein activities that are interrupted are remembered better than those that are not.

There are, of course, numerous individual differences in the basic parameters of the neuroanatomy at the routine level which translate into differences in the construction and use of routines. Any of the routine-level subsystems—perception, motor control, learning, memory—will vary in their sensitivity, and capacity for and speed of processing. These, in turn, translate into differences in the rate at which individuals can integrate information, learn skills, or acquire and recall information. Important differences for personality theorists include the sensitivity of the routine level to interruption from below (i.e., reactive level) or to control from above (i.e., reflective level; see Fig. 7.1). There might also be differences in sensitivity to sensory cues and in the tendency to do broad, global processing rather than more narrowly focused processing.

In addition, whereas reactive-level processes are essentially fixed by biology, much of the content at the routine level is learned. Because complex skills are heavily dependent on the substrate of prior learned material, individual differences in experiences and learning accumulated throughout life make for eventual large differences in abilities. Thus, both biological (genetic) and environmental (learned) differences emerge at the routine level.

Affect at the Reflective Level: Cognitively Elaborated Emotions

Reflection is a special characteristic of higher animals, most marked in primates and especially humans. Humans can construct and use mental models of the people, animals, and artifacts with which they do or could interact, as well as models of those interactions. Rich representational structures of this kind enable complex understanding, active predictions, and assessments of causal relations. Humans also have a notion of self; we have self-awareness, consciousness, and importantly, representations of the minds of others. This leads to the possibility of elaborate systems of competition and to the ability to lie and deceive, but it also leads to more sophisticated social cooperation and to a propensity for humor, art, and the like. Monkeys and apes may share some of these cognitive abilities (e.g., deWaal & Berger, 2000), but these abilities remain preeminently human.

The kinds of capability that comprise the enhanced processing of the reflective level depend on the ability of the reflective level to perceive, analyze, and in some cases, alter its own functioning as well as that of the routine and reactive levels. Humans (at least) can examine their own behaviors and mental operations, reflect upon them, and thereby enhance learning, form generalizations, predict future events, plan, problem-solve, and make decisions about what to do. In general, the reflective level comprises consciousness together with all of the advanced cognitive and metacognitive skills that have enabled humans to increase their knowledge cumulatively over the millennia.

We consider the well-established finding that prefrontal regions of the brain subserve the programming, regulation, and verification of activity (e.g., Damasio, 1994; Goldberg, 2001) as support for the separability of the kind of conscious control functions of the reflective level from other, more automatic behaviors. The fact that prefrontal damage does not affect routine behavior or the performance of well-learned skills is also consistent this view. Note that in our model—and in any model that identifies the prefrontal lobes as the locus of such activities—the reflective level neither receives direct perceptual information as input nor directly controls motor output. This means that the reflective level can only bias the levels beneath it. Norman and Shallice (1986) viewed this bias signal as "will." In their model, will is a control signal such that if some activity at a lower level is desired, the control level can add activation signals to it, thereby increasing the likelihood that it will get performed.

It is the power of the reflective level that makes possible the rich emotional experience that we assume is unique to humans. At the reflective level, not only are emotions and their associated behaviors sometimes actually initiated, as when reminiscing about prior experiences can lead to changes in moods and emotions, but less well-defined affective states become elaborated, interpreted, and transformed into full-fledged emotions. Thus, whereas at the reactive level there is only unelaborated proto-affect and at the routine level only feelings and primitive emotions, the reflective level has the capacity to interpret unelaborated proto-affect from the reactive level and primitive emotions and feelings from the routine level so as to generate discrete emotions that can be labeled. This cognitive elaboration comes about by relating higher-level cognitive representations and processes to the kind of internal and external events that induce affect in the first place.

Because the reflective level is the locus of all high-level cognitive processing, it has a rich repertoire of representational and processing resources. In addition to goals, standards, and tastes, the three classes of emotion-relevant representations identified by Ortony, Clore, and Collins (1988), these resources include such things as conscious expectations; plans; mental models and simulations; deductive, inductive, and counterfactual reasoning; and so on. At this level, it is possible to take feelings as objects of thought: we can (sometimes) label them, we try to make sense of them, and we can plan actions around them.

To illustrate this, consider the consequences of reflecting upon realized or unrealized potentials (e.g., fulfilled vs. violated expectations). The two future-oriented emotions, 3 and 4 discussed in the preceding section, have associated with them a further pair of states—one corresponding to the potential being realized (e.g., a confirmed expectation) and the other corresponding to the potential not being realized (e.g., a disconfirmed expectation). The emotions that derive from 3 (a [positive] feeling about a potential good thing) are:

3.1. A (positive) feeling about a **potential good thing, realized**
3.2. A (negative) feeling about a **potential good thing, not realized**

The emotions that derive from 4 (a negative feeling about a potential bad thing), are:

4.1. A (positive) feeling about a **potential bad thing, not realized**
4.2. A (negative) feeling about a **potential bad thing, realized**

These are four full-fledged emotional states deriving from primitive emotions or emotional feelings originally experienced at the routine level. They are affective because they involve the evaluation of something as good or bad, helpful or harmful, beneficial or dangerous, and so on; they are feelings because they inherit feeling qualities from their lower origins, albeit now changed and augmented by cognition; and they are emotions because they are about something (Clore & Ortony, 2000) and have consciously accessible content.

Of course, as anyone who has ever acted in the heat of the moment knows, strong emotions and their routine-level behavioral concomitants often overwhelm cool reason and its more planful reflective-level responses; but this very fact presupposes, rather than vitiates, the routine–reflective distinction. In fact, there are several reasons why careful, logical planning activities at the reflective level might be thwarted. One such reason is that routine-level responses might become initiated before the reflective level has completed its analysis. Another is that inhibitory signals initiated at the reflective level are too weak to overcome the automatic procedures initiated at the routine level. Finally, the emotional state might cause hormonal states that bias the reflective processes to do more shallow processing, presumably in an effort to quicken their responses, thus generating responses that are logical at the surface but that have severe negative results that would have been predicted had the reflective processes been allowed to continue. Emotional responses are often first-order responses to situations, with poor long-term impact.

It may be informative to consider an example that illustrates the rapid, automatic action at the routine level, preceding both thoughtful planning at the reflective level as well as the delayed interpretation of the resulting affective state. Many years ago, one of the authors spent a year living in a coastal town in tropical Africa. One day, on his way to the beach, he was driving slowly and with considerable difficulty across a shallow, rough, dried-up riverbed with his car windows open. Suddenly, and quite unexpectedly, he saw a huge crocodile that had been lying still on the riverbed, now disturbed by the approaching car. Panicked, he put his foot on the brake pedal to stop the car, leaned across the unoccupied passenger seat, and frantically rolled up the window on the side where the crocodile was. Having done this, he rolled up the window on his (driver's) side and, shaking and heart pounding, drove, still slowly and with difficulty, out of the riverbed, to what he took to be safety. Then, and only then, did he become aware of how terrified he was.

In this example, a potential threat was perceived and a rapid protective-behavior routine initiated. There was too little time to optimize the selected routine. The system was satisficing rather than optimizing. Realistically, it might have made more sense to just keep going—the crocodile was not likely to climb into a moving car through the passenger door window and devour the driver. Presumably, the driver stopped the car to facilitate the closing of the window, but this was not thought through or planned—it was just done—a sequence of the "car-stopping" routine followed by the "window-closing" routine. Furthermore, the behavior is not well described by saying that it was done in response to, or even as part of, fear. As described, the emotion of fear came only after the driver had engaged in the protective behavior and extricated himself from the situation—only then, on reviewing his racing heart, his panicky and imperfect behavioral reactions, and the situation he had just been in, did he realize how frightened he was. In other words, the emotion was identified (labeled) as fear only after the behavior and concomitant feelings (of bodily changes) had been interpreted and augmented by cognition at the reflective level. The situation is best described by saying that first came the feeling of primitive fear (which includes an awareness of the bodily changes) and then, upon interpretation and additional cognitions, came the full-fledged emotion of fear.

This example not only bears upon several aspects of our three-level model but also speaks to the James-Lange theory of emotions (James, 1884; Lange, 1895/1912), especially with respect to the temporal relationship between emotions and behavior. In our example, the rapid behavior occurred before the emotion was identified, exactly as William James described it with respect to his imaginary bear in the woods:

> the bodily changes follow directly the perception of the exciting fact, and [that] our feeling of the same changes as they occur is the emotion. Common sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. The hypothesis here to be defended says that this order of sequence is incorrect . . . and that the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble . . . Without the bodily states fol-

> lowing on the perception, the latter would be purely cognitive in form, pale, colorless, destitute of emotional warmth. We might then see the bear, and judge it best to run, receive the insult and deem it right to strike, but we should not actually feel afraid or angry.

Now consider James' example of the emotion that accompanies one's loss of a fortune. In this case, it would seem that the reflective-level analyses come first. The person would start thinking about possible causes of the loss, perhaps reviewing past actions by (formerly) trusted associates and then assessing blame. Such cognitions would be likely to invoke evaluation as a result, for example, of running through various "what-if" scenarios and imagining the responses of family, friends, and colleagues. This kind of cognitively induced introduction of sources of value would be the wellspring of bodily changes, the awareness of which would constitute the underlying emotional feeling. However, if all of this were to lead to anger, the anger would have followed the cognitions. Similarly, James' emotion of "shame" results from self-blame, and this means that it is cognition, not behavior, that is the trigger. All this suggests to us that the question is not whether the James-Lange theory is right or wrong but, assuming that it is at least in part right, under what conditions it is right and under what conditions it is wrong. So, if one asks the question "Which comes first, cognition or behavior?" the answer has to be that it depends. When reactions are triggered from the reactive or routine level, behavior precedes; but when the triggering comes from the reflective level, cognition precedes.

Much as with the routine level, there are many sources of individual differences in the operating parameters of the reflective level. These are likely to include such things as sensitivity, capacity, and processing speed plus the ability of the reflective level to influence lower levels through its control signals of activation and inhibition. We would also expect to find differences in conscious working memory and attentional focus, especially with respect to sensitivity to interruptions and other events. Finally, there will be substantial individual differences in the content at both the behavioral and reflective levels, and inasmuch as the reflective level is the locus of one's self image and much cultural knowledge and self-examination, these differences can be expected to have a significant effect on the way a person interacts with the environment and with others.

IMPLICATIONS FOR PERSONALITY

We have already suggested a number of parameters for which we might expect inter- and intra-individual differences at the different levels of processing. We view parameters of this kind as the foundations of personality. Inevitable variations in parameter values lead to individuals differing in the ways in which, and the effectiveness with which, they function in the world. However, personality research lacks a consensual account of what personality is (especially with respect to its causal status), so we start our discussion by situating our account in relation to the principal current approaches to personality theory.

Most current research in personality focuses on individual differences in affect and interpersonal behavior while adopting one of two different and largely incompatible perspectives. One of these seeks to identify the primary dimensions in terms of which descriptions of systematic regularities and differences across different times and different places can be parsimoniously but informatively cast. The other perspective views personality as a causal factor in the functioning of individuals and thus seeks to identify deeper explanations of such similarities and differences. We believe that our approach can resolve some of the conflict between these two perspectives and that it moves beyond both by extending the purview of personality theory from affect and interpersonal behavior to include behavior more generally as well as motivation and cognition. For us, personality is a self-tunable system comprised of the temporal patterning of affect, motivation, cognition, and behavior. Personality states and traits (e.g., for anxiety) are a reflection of the various parameter settings that govern the functioning of the different domains at the different levels.

One of the most paradoxical yet profound characterizations of personality is the idea that all people are the same, some people are the same, and no people are the same (Kluckholm & Murray, 1953). In our terms, all people are the same in that everyone is describable in terms of the four domains of functioning (affect, motivation, cognition, and behavior) at the three levels of processing (reactive, routine, and reflective); some people are the same in that they are similar in the way that they function in some or all of the domains; and finally, no one is the same in the unique details of the way in which the four domains interact with each other and at the three processing levels.

With respect to our levels of processing, it is clear that individual differences occur at all three levels. We have already suggested possible dimensions of variability at the different levels. For example, at the reactive level one might expect differences in sensitivity to environmental stimuli, aspects of response strength, and ability to sustain responses. Such differences would manifest themselves as variations in the likelihood of approach and avoidance and in proto-affective responses (Schneirla, 1959). As outside observers, we might characterize some of these as variations in a behavioral trait. For example, one might map observed differences in probabilities of approach and avoidance onto a boldness–shyness dimension, as do Coleman and Wilson (1998) in their description of pumpkinseed sunfish.[7] More generally, individual differences at this level were discussed long ago by Pavlov and later by others in terms of strength and lability of the nervous system (Pavlov, 1930; Nebylitsyn & Gray, 1972; Robinson, 1996, 2001; Strelau, 1985).

At the routine level, individual differences become more nuanced. Consider an individual who, relative to others, has a high level of positive affect and a high likelihood of approach behaviors, both emanating from the joint effects of reactive- and routine-level processing.[8] This combination of operating parameters is typical of the trait "extraversion." In other words, the descriptive label "extravert" is applied to someone who is high on both the affective and behavioral dimensions. This additive structure will, of course, result in correlations of extraversion with positive affect and with approach behavior but not necessarily to high correlations between responses across the different domains (i.e., of positive affect with approach behaviors). Our view is that the reason that we call someone an extravert is that they tend to do things such as go to lively parties (behavior) and they tend to be happy (affect). Similarly, the descriptive term for an emotionally less stable individual ("neurotic") reflects a larger likelihood of negative affect as well as a higher likelihood of avoidance behaviors. Although many situations that induce negative affect also induce avoidance behaviors, and thus make individual differences in negative affect and avoidance more salient, "neuroticism" is merely the label applied to those who are particularly likely to experience high negative affect while avoiding potentially threatening situations. (A somewhat similar argument was made by Watson, 2000, who emphasized the affective nature of extraversion and neuroticism and considered the functional nature of approach and withdrawal behavior in eliciting affect.) The virtue of this account is that it explains the fact that reliably large correlations across domains of functioning are hard to find. From the point of view of the parameters that control their operation, the domains of functioning are largely independent.

Although there are exceptions, most personality inventories and rating scales are designed to get at what we consider to be routine-level activity (although they do so by soliciting reflective-level responses). Such measures often use items that tap separately the different domains. Thus, an item like "Do you feel nervous in the presence of others?" is an attempt to get at routine-level affect, the item "Do you avoid meeting new people?" addresses routine-level behavior, and the item "Does your mind often wander when taking a test?" addresses routine-level cognition. To be sure, someone who is high on all three of these items is likely to act and feel very differently from someone who is low on all three. However, because for each person the parameter settings in the different domains of functioning are probably independent, a value on one item (domain) does not predict the value of any others.

At the reflective level, we see the complex interplay of individual differences in motivational structures (e.g., promotion and prevention focus; Higgins, 2000) with cognitive representations (e.g., attributions of success and failure; Elliot & Thrash, 2002) that lead to the complex affective and behavioral responses we think of as effective functioning. It is also at this level that people organize life stories to explain to themselves and others why they have made particular life choices (McAdams, 2001).

We suspect that most, if not all, of the five major domains of the traditional descriptive approach to personality (see John & Srivastava, 1999, for a discussion) can be accounted for by individual differences in the parameters and content of the three levels of processing and the four domains of functioning. As we have already discussed, differences at the reactive level reflect differences in sensitivities to environmental situations. The reactive level is probably also the home of phobias such as fear of heights, crowds, darkness, snakes, spiders, and so on, which might explain why these are relatively easy to acquire but very difficult to extinguish. Routine- and reflective-level differences will exist both at the biological substrate and in learned routines, behavioral strategies, and cultural norms. These will probably determine many of the "Big 5" parameters, with neuroticism and extraversion and parts of agreeableness and conscientiousness probably due to routine-level differences and openness and the more planful parts of conscientiousness due to more reflective-level concerns (see also Arkin's Chapter 9).

By conceptualizing personality in terms of levels of processing and domains of functioning, we believe that we can improve upon prior personality research that has tended to focus on functioning drawn from only one domain at a time (e.g., affect and neuroticism or approach behavior and extraversion). We also think that by applying this approach we will be able to integrate biologically and causally oriented theories with descriptive taxonomies, which, while perhaps lacking explanatory power, have nevertheless been quite useful in predicting functioning in real-life settings (e.g., job performance in the workplace; Barrick & Mount, 1991).

IMPLICATIONS FOR THE DESIGN OF AUTONOMOUS ROBOTS AND OTHER COMPLEX COMPUTATIONAL ARTIFACTS

In animals, affect, motivation, cognition, and behavior are all intertwined as part of an effective functioning system. There is no reason to believe that it should be any different for intelligent, socialized robots and autonomous agents, physical or virtual. Just as species at different levels of evolutionary complexity differ in their affective and cognitive abilities, so too will different machines differ. A simple artifact, such as a robotic vacuum cleaner, is implemented as a purely reactive-level device. At this level, affect, motivation, and behavior cannot be separated from one another. Such a device has the analog of hard-wired drives and associated goal states. When there is conflict, it can be resolved by the kind of subsumption architecture described by Brooks, which has been implemented in a variety of simple robots (e.g., Brooks, 1986, 2002; see Chapter 10).

More complex artifacts that can perform large numbers of complex tasks under a variety of constraints require routine-level competence. Thus, SOAR, the cognitive modeling system that learns expert skills, is primarily a routine-level system (Rosenbloom, Laird, & Newell, 1993). In fact, expert systems are quintessentially routine-level systems. They are quite capable of expert performance but only within their domain of excellence. They lack higher-level monitoring of ongoing processes and extra-domain supervisory processes. Finally, when HAL, the fictional computer in the movie *2001*, says "I'm afraid, Dave," it is clearly identifiable as a reflective-level computational artifact (assuming that the statement resulted from consideration of its own state). Whether any artifact today operates at the reflective level is doubtful. To address the question of what it would take for this to happen, we now examine how the model of effective functioning that we have sketched might apply to autonomous robots and other complex computational artifacts. In doing so, we will pay special attention to the functional utility of affect for an organism, be it real or synthetic.

We believe that our model, which integrates reactive- and routine-level processing with reflective-level processing and incorporates the crucial functions played by affect, constitutes a good way of thinking about the design of computational artifacts. This is particularly so for artifacts of arbitrary complexity that must perform unanticipated tasks in unpredictable environments. When the task and environment are highly constrained and predictable, it is always appropriate and usually possible to use strong methods (Newell & Simon, 1972) and build a special-purpose device that performs efficiently and successfully, as is current practice with most of today's industrial robots. However, under less constrained tasks and environments, strong methods are inadequate unless the system is capable of producing new mechanisms for itself. A system capable of generating its own, new, special-purpose mechanisms would necessarily employ some weak methods and would probably need an architecture of similar complexity to the one we are proposing.

Implications of the Processing Levels

In the early days of artificial intelligence and cognitive psychology, considerable attention was devoted to how to best represent general and specific knowledge, plans, goals, and other cognitive constructs and how to do higher-order cognitive functioning such as language understanding, problem solving, categorization, and concept formation. To some extent, this ignored motivation, which, of course, is necessary to explain why an organism would establish a goal or develop a plan in the first place. Ironically, behaviorist psychologists—the very people against whom the cognitivists were reacting—had worried about these issues and had even proposed biologically plausible models of the causes of action initiation (e.g., the dynamics of action model; Atkinson & Birch, 1970). We think that recent revivals of this model (e.g., Revelle, 1986) can do a reasonable job of accounting for a good deal of action initiation at our reactive and routine levels.

It is easy to understand why a robot—or any organism, for that matter—acts when confronted with environmental conditions (or internal drives) that demand some kind of response; but what happens when they are not imposing any demands on the organism and it is at or close to homeostasis? Does it then just remain idle until some new action-demanding condition arises that causes it to behave? Animals' motivation systems handle this by letting the resting point of affect be slightly positive so that when there is nothing that needs to be done, the animal is led to explore the environment (see Cacioppo, Gardner, & Berntson, 1997, on positivity offset). This is the affective basis of *curiosity* (an innate motivational force that leads organisms to explore the environment and to try new things). Certainly in humans, curiosity (openness to experience) is a powerful learning aid. So should it be for a robot. Clearly, an autonomous robot is going to need expectations. Perceiving and acting in the world while indifferent to outcomes would hardly be conducive to effective functioning. At the routine level, our model provides implicit expectations (in contrast to the conscious expectations and predictions of the reflective level). Expectations are important not only because their confirmations and disconfirmations are crucial for learning but also because the resulting affect changes the operating characteristics of the other three domains. At the routine level, implicit expectations are tightly bound to their associated routines. They come into play much less often when routines run off successfully than when they fail or are interrupted. Recall that at this level proto-affect from the reactive level becomes partially elaborated as primitive emotions (feeling good or bad about the present or potential future). In designing an autonomous robot, we would need to consider the motivational, cognitive, and behavioral consequences of these primitive emotions. Consider the simplest case, that of feeling good or bad about the present. Part of the power of affective states in general derives from the fact that they are the result of mapping many inputs onto a few or, in the limiting case, two (positive and negative) kinds of internal state. For example, any of a multitude of disconfirmed positive expectations or confirmed negative ones can reduce one to the primitive emotional state that we might call displeasure or distress. This affective state in turn functions as a simple modulator of processing parameters in the other three domains of functioning. Thus, the power of affect, and hence its value for robot design, is its data-reduction capacity and consequent parameter-modulating properties. In animals, the magnitude and even direction of changes that result from an affective state vary from individual to individual and comprise an important part of personality. We would expect to include the potential for such differences in the design of automata.

Finally, we need to consider the implications of adding reflective-level capacities. To do this, we have to enable the robot to have active expectations about outcomes and states of the world. In addition, it will have to be able to reflect on its own actions and states, a capacity that is critical for the formation of generalized knowledge, for abstraction, and for developing principles and new knowledge representations. Some of these representations (e.g., plans, goals, standards, and values) are themselves unique inhabitants of the reflective level, providing the basis for more fine-grained appraisals of emotion-inducing events and the material necessary for interpreting feelings as emotions.

Affect and Emotion

As soon as one raises the topic of affect and emotion in artifacts, one has to confront the probably unanswerable philosophical question of whether robots can have feelings (see Chapter 2). We choose to finesse this question by restricting our attention to the functional utility of affect and emotion. We view feeling as an awareness of a bodily state, a bodily disturbance, or some other bodily change. However, neither we nor anyone else know how to incorporate the experience of such awareness into an inanimate artifact.

With respect to the functional utility of affect, consider first the value of emotion recognition, a crucial capacity for the social aspect of effective functioning. Effective social functioning involves understanding the conditions under which it is or is not appropriate or prudent to interact with other individuals and when it is deemed appropriate, knowing what kind of interaction is expected. However, this ability to recognize, understand, and predict the current affective state of others, *emotional intelligence* (e.g., Mayer & Salovey, 1997), is not the only determinant of effective social functioning. It is also advantageous to be able to make inferences from a model of the longer-term patterns of affective and motivational states, cognitions, and behavior—that is, from a model of the individual's personality. For example, our reflective characterization of a person as momentarily happy or sad and dispositionally moody or hyperactive contributes to the decisions we might make about our actions and interactions with respect to that person. Thus, a socially savvy robot will need to make inferences from behavior and outward manifestations of emotions (emotional expression), motivations, and cognition as well as from its model of the personality of others, when available, if it is to be capable of effective social functioning.

So, there are good reasons why a robot might need to recognize affect in others; now we need to ask why it might need affect itself. Our answer is that robots need affect for the same reason that humans do. One of the most fundamental functions of affect is as a valenced index of importance, and indeed, there is some neuroscientific evidence that affect is a prerequisite for establishing long-term memories (e.g., McGaugh, Cahill, Ferry, & Roozendaal, 2000). A second important function of affect is that it provides occasions for learning, from quite simple forms of reinforcement learning to complex, conscious planning and experimentation. Affect also has important consequences for the allocation of attention. It is a well-established finding in the psychological literature that negative affect tends to result in the focusing of attention on local details at the expense of global structure. Presumably, this is because in times of stress or threat it is important to be vigilant and to attend to local details, to identify sources of potential danger. Focusing attention on large-scale, global conditions of the environment is not likely to be conducive to these goals. However, such global focus is likely to be valuable in situations that are devoid of threat, danger, or potential harm. Consistent with this idea is the fact that under conditions of positive affect people do tend to focus on the big picture and to engage in more expansive information processing (Ashby, Isen, & Turken, 1999; Gasper & Clore, 2002). All of these (and indeed other) functions of affect are achieved through its capacity to change the operating characteristics of the other domains of functioning—motivation, cognition, and behavior. For example, the negative affect that results from the perception of a threat might modulate motivation by increasing the strength of a self-protecting action tendency, such as running away, relative to, say, an enjoyment-seeking action tendency, such as having a cocktail. Similarly, the affect might modulate cognition by interrupting ongoing cognitive processes and focusing attention on details of the current problem, and of course, it is almost bound to change the ongoing behavior.

CONCLUSION

We have presented a general model of effective functioning conceptualized in terms of three levels of processing (Fig. 7.1), in which four domains of functioning (affect, motivation, cognition, and behavior) are seen as integrated, nonseparable components. The reactive level is the home of rapid detection of states of the world and immediate responses to them. It uses pattern matching to recognize a set of situations and stimuli for which it is biologically prepared. These are essentially the unconditioned stimuli and associated responses of the simplest forms of classical conditioning. The reactive level is essential for mobilizing appropriate responses to the exigencies of the environment, and it can interrupt higher-level processing. The routine level is that of most motor behavior as well as procedural knowledge and automatic skills. It is a complex, rich information-processing and control system. It too interrupts higher-level processing when it encounters unexpected conditions, impasses, or emergencies or when conditions are novel or unknown. The reflective level is that of conscious attention, of higher-level cognitive processes and representations, and of cognitively elaborated, full-blown emotions. It is also the home of reflection and of knowledge about one's own knowledge and behavior. As such, this system continually performs high-level monitoring of ongoing activity at all three levels. The reflective level does not receive direct sensory input nor does it directly control responses: it can only potentiate or inhibit activity at the lower levels.

Within this three-level architecture, we have considered the way in which the four domains of functioning interact, with special attention to the way in which affect is manifested at the different levels. In many respects, labeling these continuous, complex feedback systems in terms of the four common distinctions of affect, motivation, cognition, and behavior is somewhat arbitrary. This is an integrated, holistic system that has evolved to facilitate effective functioning in a complex, dynamic environment. Nature does not necessarily make the sharp distinctions among these levels and domains that we make in order to talk about them. Affect, for example, ranges from proto-affect at the reactive level through primitive emotions at the routine level to full-blown emotions when augmented with the other domains at the reflective level. Thus, full-fledged emotions can involve feelings from the somatic and motor components of the reactive level, interacting with proto-affect from the reactive level and primitive emotions and feelings from the routine level together with cognitive elaboration from the reflective level. Reflective affect without some contribution from lower levels cannot be full-blown, "hot" emotion. For example, the cognitive components of anger without the concomitant feeling components from the lower levels would be what we might call "cold, rational anger." Similarly, the feeling of primitive fear at the routine level is not a full-blown emotion because it lacks the requisite cognitive elaboration. It is only a feeling (albeit unpleasant) waiting to be "made sense of" by reflective-level processes.

As we indicated at the outset, the model that we have proposed is best thought of as a framework for thinking about how to conceptualize effective functioning. We believe that it is only by considering functioning at all three levels of processing and at all four domains of functioning that we can expect to achieve an understanding of effective functioning that might be useful for the design of fully autonomous robots and agents capable of responding appropriately to the huge array of problems and prospects that their environments might present.

Notes

We thank Ian Horswill, for his helpful comments on an early draft of this chapter, and Tony Z. Tang, for helpful discussions in the early stages of this work.

1. Although some investigators view cognition as a form of behavior (e.g., Fellous, 1999), we prefer to make a sharp distinction between the two.

2. We are well aware that talking of learning at this level simply in terms of classical conditioning is far too simplistic. Razran (1971) provides a brilliant discussion of the complexities of this issue.

3. Following Watson and Tellegen (1985) and others, we view positive affect and negative affect as (at least partially) independent dimensions.

4. Note that although initially bitter tastes are rejected, the system can adapt so that, with sufficient experience, it no longer responds quite so vehemently. Indeed, the higher levels might interpret the taste positively and actively inhibit the lower response—hence, the learned preference for many bitter and otherwise initially rejected foods such as alcoholic beverages and spicy sauces.

5. The key feature of specifying emotions (and emotion-like states, such as 1–4) in this way is that they are characterized in terms of their eliciting conditions with minimal dependence on the use of emotion words. The advantage of doing this can be seen by considering that English does not have a good word to express the affective state characterized by 3, a positive feeling about a potential good thing. Something like "anticipatory excitement" is much closer to the state than "hope," even though in English hope is usually opposed to fear. In any case, we think it misleading to use conventional emotion names to refer to primitive forms of emotion.

6. By "internal" cue to the reactive level we mean internal to the organism but still external to the reactive-level mechanisms. Thus, in the case of hunger, the internal cue to the reactive level comes from the hunger system.

7. In fact, our preference would be to view boldness and shyness as two independent, unipolar dimensions rather than one bipolar dimension. We also suspect that *timidity* is a better term to capture the construct because it avoids the social connotations of "shyness."

8. Having a high level of positive affect does not mean that the individual is always happy. It means that the median value of positive affective responses is higher for this individual than for most others. The same is true for approach behaviors (indeed, for everything).

References

Arnold, M. (1960). *Emotions and personality* (Vols. I, II). New York: Columbia University Press.

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529–550.

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. New York: Wiley.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance. *Personnel Psychology, 44*, 1–26.

Broadbent, D. E. (1971). *Decision and stress*. London: Academic Press.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, RA-2*, 14–23.

Brooks, R. A. (2002). *Flesh and machines: How robots will change us*. New York: Pantheon.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3–25.

Clore, G. L., & Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never? In L. Nadel, R. Lane, & G. L. Ahern (Eds.), *The cognitive neuroscience of emotion*. New York: Oxford University Press.

Coleman, K., & Wilson, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: Individual differences are context-specific. *Animal Behaviour, 56*, 927–936.

Damasio, A. (1994). *Descarte's error: Emotion, reason, and the human brain*. New York: Putnam.

Damasio, A. R. (2000). A second chance for emotion. In L. Nadel, R. Lane, & G. L. Ahern (Eds.), *The cognitive neuroscience of emotion*. New York: Oxford University Press.

deWaal, F. B. M., & Berger, M. L. (2000). Payment for labour in monkeys. *Nature, 404*, 563.

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

Fellous, J.-M. (1999). The neuromodulatory basis of emotion. *The Neuroscientist, 5*, 283–294.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science, 13*, 34–40.

Goldberg, E. (2001). *The executive brain: Frontal lobes and the civilized mind*. New York: Oxford University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion, 4,* 269–288.

Halliday, T. R. (1980). Motivational systems and interactions between activities. In F. Toates & T. R. Halliday (Eds.), *Analysis of motivational processes* (pp. 205–220). London: Academic Press.

Higgins, E. T. (2000). Does personality provide unique explanations for behavior? Personality as cross-person variability in general principles. *European Journal of Personality, 14,* 391–406.

James, W. (1884). What is an emotion? *Mind, 9,* 188–205.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York: Guilford.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to alternatives. *Psychological Review, 93,* 136–153.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A., & Whalen, P. J. (2004). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 14,* 2317–2322.

Kluckholm, C., & Murray, H. (1953). *Personality in nature, society, and culture.* New York: Knopf.

Lane, R. (2000). Neural correlates of conscious emotional experience. In L. Nadel, R. Lane, & G. L. Ahern (Eds.). *The cognitive neuroscience of emotion.* New York: Oxford University Press.

Lange, G. (1912). The mechanism of the emotions. In B. Rand (Ed., Trans.), *The classical psychologists: Selections illustrating psychology from Anaxagoras to Wundt* (pp. 672–684). Boston: Houghton Mifflin. (Original work published 1895)

Lazarus, R. S. (1966). *Psychological stress and the coping process.* New York: McGraw-Hill.

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

MacLean, P. D. (1990). *The triune brain in evolution*: Plenum.

Mandler, G. (1984). *Mind and body.* New York: Wiley.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3–31). New York: Basic Books.

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5,* 100–122.

McGaugh, J. L., Cahill, L., Ferry, B., & Roozendaal, R. (2000). Brain systems and the regulation of memory consolidation. In J. J. Bolhuis (Ed.), *Brain, perception, memory: Advances in cognitive neuroscience* (pp. 233–251). London: Oxford University Press.

Minsky, M. (in preparation). *The emotion machine.*

Nebylitsyn, V. D., & Gray, J. A. (1972). *The biological basis of individual behavior.* New York: Academic Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation: Advances in research* (Vol. IV). New York: Plenum.

Öhman, A., Flykt, A., & Lundqvist, A. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In L. Nadel, R. Lane, & G. L. Ahern (Eds.), *The cognitive neuroscience of emotion*. New York: Oxford University Press.

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

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science, 11*, 341–364.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315–331.

Pavlov, I. P. (1930). A brief outline of the higher nervous activity. In C. A. Murchinson (Ed.), *Psychologies of 1930*, Worchester, MA: Clark University Press.

Razran, G. (1971). *Mind in evolution*. Boston: Houghton Mifflin.

Revelle, W. (1986). Motivation and efficiency of cognitive performance. In D. R. Brown & J. Veroff (Eds.), *Frontiers of motivational psychology: Essays in honor of John W. Atkinson*. Berlin: Springer.

Revelle, W. (1993). Individual differences in personality and motivation: "Non-cognitive" determinants of cognitive performance. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control. A tribute to Donald Broadbent*. Oxford: Oxford University Press.

Robinson, D. L. (1996). *Brain, mind, and behavior: A new perspective on human nature*. London: Praeger.

Robinson, D. L. (2001). How brain arousal systems determine different temperament types and the major dimensions of personality. *Personality and Individual Differences, 31*, 1233–1259.

Rolls, E. T. (1999). *The brain and emotion*. New York: Oxford University Press.

Roseman, I. J. (1984). Cognitive determinants of emotions: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 5). Beverly Hills: Sage.

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1993). *The SOAR papers*. Boston: MIT Press.

Sanders, A. F. (1986). Energetical states underlying task performance. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 139–154). Dordrecht: Matinus Nijhoff.

Scherer, K. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer and P. Ekman (Eds.), *Approaches to emotion*. Hillsdale, NJ: Erlbaum.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schneirla, T. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In *Nebraska symposium on motivation* (pp. 27–58). Lincoln: University of Nebraska Press.

Sloman, A., & Logan, B. (2000). Evolvable architectures for human-like minds. In G. Hatano, N. Okada, & H. Tanabe (Eds.), *Affective minds* (pp. 169–181). Amsterdam: Elsevier.

Strelau, J. (1985). Temperament and personality: Pavlov and beyond. In J. Strealu, F. H. Farley, & A. Gale (Eds.), *The biological bases of personality and behavior: Psychophysiology, performance, and application*. Washington, DC: Hemisphere.

Watson, D. (2000). *Mood and temperament*. New York: Guilford.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–235.

Zeigarnik, B. (1967). On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Humanities. (Original work published 1927)