Applied Artificial Intelligence, 16:577–609, 2002

0883-9514/02 $12.00 + .00
DOI: 10.1080/0883951029003040 8

CAN A RATIONAL AGENT AFFORD TO BE AFFECTLESS? A FORMAL APPROACH

CHRISTINE LÆTITIA LISETTI
Department of Computer Science,
University of Central Florida, Orlando, Florida, USA

PIOTR GMYTRASIEWICZ
Department of Computer Science,
University of Illinois at Chicago, Chicago, Illinois, USA

In this article, we expose some of the issues raised by the critics of the neoclassical approach to rational agent modeling and we propose a formal approach for the design of artificial rational agents that includes some of the functions of emotions found in the human system. We suggest that emotions and rationality are closely linked in the human mind (and in the body, for that matter) and, therefore, need to be included in architectures for designing rational artificial agents, whether these agents are to interact with humans, to model humans' behaviors and actions, or both.

We describe an Affective Knowledge Representation (AKR) scheme to represent emotion schemata, which we developed to guide the design of a variety of socially intelligent artificial agents. Our approach focuses on the notion of ''social expertise'' of socially intelligent agents in terms of their external behavior and internal motivational goal-based abilities. AKR, which uses probabilistic frames, is derived from combining multiple emotion theories into a hierarchical model of affective phenomena useful for artificial agent design. AKR includes a taxonomy of affect, mood, emotion, and personality, and a framework for emotional state dynamics using probabilistic Markov Models.

INTRODUCTION

Until recently, most modern scientific theories followed the ''Cartesian Mission'' of the seventeenth century, which worked hard to establish indubitable objective truths about the world ''out there.'' Placing reason on a

Dr. Lisetti would like thank Dr. David Rumelhart and Dr. Brian Arthur for inviting her to attend the 1996 *Workshop on Economics and Cognition* at the Santa Fe Institute when many of the ideas exposed in this paper took shape. She also would like to acknowledge that her research is supported, in part, by the Office of Naval Research Prime Award 2108-206L0.

Address correspondence to C. Lisetti, Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA. E-mail: lisetti@cs.ucf.edu

pedestal (with thinkers such as Descartes, Newton, and others of their contemporaries), rationalism was indeed necessary in order to allow science to build itself up against the old generation of mythical and mystical thought, turning its back on the "delusive" world of senses and perception, and relying on mathematical properties of the real world which could be trusted. From the rationalist perspective, the human intellectual mind was a reflection of the real world, a world of mathematical properties at odds with the false testimony of the senses. Reason and intellect, in the Cartesian tradition, was opposed to emotions and feelings, which undermined the latter, with the disruptive "passions of the soul."

Since then, however, many scientists have become increasingly interested in notions, such as chaos in physics (which looks at the other side of order), emotions, and the unconscious in psychology, sociology, and cognitive science (which question human rationality in its classical sense) (Tomkins and Izard 1966; Arnold 1960; Plutchik 1980; Izard 1977; Ekman et al. 1971; Leventhal and Tomarken 1986; Frijda 1986; Ekman et al. 1994; Zajonc 1980; Frijda 1986; de Sousa 1990; Damasio 1994; Lutz 1985; Ortony et al. 1988; Collins 1975; Barbalet 1998); metaphors in linguistics (which can shape human thought) (Lakoff 1987); cognitive effects of emergent neurological patterns in philosophy, artificial intelligence (AI, henceforth), and neuroscience (which strongly questions the Cartesian three-century-old "mind-body problem") (Rorty 1979; Rumelhart and McClelland 1986); the effects of cognition on the body proper in medicine and Eastern philosophy; the influence of individuals' beliefs in shaping the macro-economy in economics (which relates utility with *dynamic* individuals' preferences) (Arthur 1995), situated cognition in anthropology, AI, and robotics (which looks at cognition as emerging from a continuous *interaction* between the cognizing and its environment) (Clancey 1997; Hutchins 1995; Varela 1989; Varela et al. 1991; Brooks 1987; Clark 1996); and affective computing in computer science which studies "computing that relates to, arises from, or deliberately influences emotions" (Picard 1997).

Some scientists in these disciplines are working to reframe the questions raised by the scientific community and account for an endogeneous understanding of reality, in contrast with Cartesian-Newtonian science concerned with finding objective properties of reality. This shift of attention brings about a focus on the irregular side of nature—the discontinuous and the erratic, as well as the unconscious side of human nature. Connections between the different kinds of irregularities are being sought by turning to a science of chaos, i.e., a science of process rather than state.

In this article, we focus on the importance of emotional intelligence for the overall human performance in tasks, such as rational decision making, communicating, negotiating, and adapting to unpredictable environments. As a result, we claim that people can no longer be modeled as pure goal-driven, task-solving agents: they also have emotive reasons for their choices

and behavior, which (more often than not, we will discuss) participate in their rational decision making (Mandler 1975).

We expose some of the issues raised by the critics of the neo classical approach to rational agent modeling and propose a formal approach for the design of artificial rational agents that includes some of the functions of emotions found in the human system. We suggest that emotions and rationality are closely linked in the human mind (and in the body, for that matter) (Korzybski 1933) and, therefore, need to be included in architectures for designing rational artificial agents, whether these agents are to interact with humans, to model humans' behaviors and actions, or both.[1]

We describe an Affective Knowledge Representation (AKR) scheme to represent emotion schemata, which we developed to guide the design of a variety of socially intelligent artificial agents. Our approach focuses on the notion of "social expertise" of socially intelligent agents in terms of their external behavior and internal motivational goal-based abilities. AKR, which uses probabilistic frames, is derived from combining multiple emotion theories into a hierarchical model of affective phenomena useful for artificial agent design. AKR includes a taxonomy of affect, mood, emotion, and personality, and a framework for emotional state dynamics using probabilistic Markov Models.

NORMATIVE RATIONAL AGENT MODELING

Normative theories (different from descriptive psychological ones) have been proposed to account for rationality in terms of both reasoning and decision making: formal logic and decision theory, respectively (Johnson-Laird and Shafir 1993). Since we are interested in the ways in which humans make decisions and how affect enters the rational framework (which we will argue in a later section), we are not interested in reasoning which, busy with drawing valid inferences through the rules of logic, does not interact with affect.

Rational decision making, on the other hand, is highly influenced by affect and models of such processes need to include its role. In order to set our framework, we first describe the neoclassical approach to rational agent modeling, its assumptions, and advantages. We will later show how it can be enhanced and modified to include the impact of affect in rational decisions.

Normative neo classical theory of agent modeling is based on decision theory (DT, henceforth), itself based upon the notion of *choice* as determining actions and pictured in Figure 1. In short, a rational agent is one who, given its beliefs about the environment, chooses an action in such a way as to maximize its desires or subjective expected utility (SEU). SEU is an expression of the subjective probability of an outcome associated with a choice, and of its utility in terms of the value or pay-off to the decision maker. The beliefs

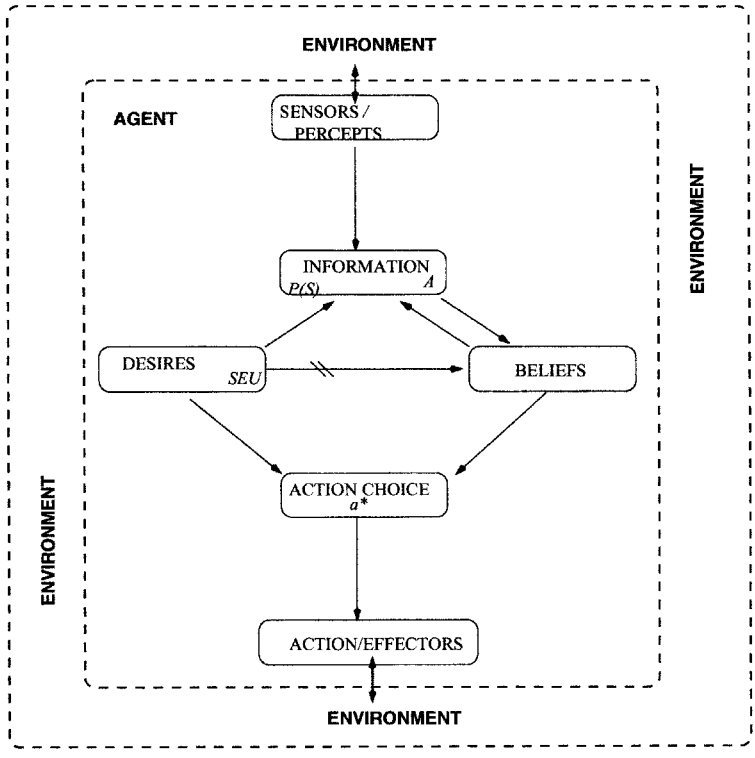

FIGURE 1. Rational utility-based agent.

themselves must be optimal, that is, unbiased by the "hot" mistakes caused by motivational biases, shown with the blocked arrow from desires to beliefs in Figure 1.

Axioms of the Normative Rationalist Model

Following the *Logical Positivist* tradition—which holds that all knowledge can be characterized by logical theories connected to observation sentences themselves corresponding to sensory inputs—some axioms of the neoclassical theory can be formulated and considered either as truths or as assumed truths (Lane and Maxfield 1995).

Short of some more technical points, the axioms of the theory can be encapsulated as follows:

1. At any point time, an agent is in some *state of the world* $s \in S$, where S is the set of all possible states of the world.
2. Because the entire world is not always known nor observable, each state of the world is associated with *probability distributions over the states of*

the world, $\mathbf{P}(S)$, quantifying how likely each of these states are (see Figure 1).

3. One of these distributions, say $P_c(S)(\in \mathbf{P})$, specifies which of these states are currently possible and how likely they are. Thus $P_c(S)$ fully describes the information the agent has about the *present state of the world* (or the environment).
4. Every action an agent takes is the result of a *choice*.
5. In order to choose an action, an agent must:
   - Decide upon a finite set of available acts, the *action space* A, containing all of the possible distinct actions a_i that the agent can take at time t.
   - Associate *consequences with each potential action* $a_i \in A$:
     - for each action, identify all possible *resulting states*;
     - project the *likelihood of each possible resulting state* using a function,
       Proj: $\mathbf{P}(S) \times A \rightarrow \mathbf{P}(S)$, which gives another probability distribution $P_i(S)(\in \mathbf{P})$, also over S.

In addition, in neo classical theories, agents are assumed to be *perfectly rational* beings. Rationality imposes constraints upon the kinds of behavior an agent can exhibit. One way to ensure the notion of rationality is to impose an axiomatized table of the result of each choice (the possible resulting states of the world) associated with *objective values* (i.e. real numbers). In other words, rational agents are considered to obey some external and context-independent preference order. Hence, another axiom can be expressed as follows to address the notion of rationality:

6. An agent's *rationality* is insured by assuming that the agent must choose an action depending on a predetermined value of the consequences associated with the action calculated with a *utility function* $U : S \rightarrow R$, such that it chooses the action a^*, that *maximizes* the expected utility of the result:

$$a^* = \underset{a_i \in A}{\text{ArgMax}} \sum_{s^j \in S} p_i^j U(s^j), \tag{1}$$

where the p_i^j is the probability that the projected distribution $P_i(S)$ assigns to a state $s^j \in S$.

The elements defined above are sufficient to formally define the *rational decision-making situations* of an agent:

Decision-making situation definition. A decision-making situation of an agent is a quadruple: $D = \langle P_c(S), A, Proj, U \rangle$, where $S, P_c(S)$, A, *Proj* and U are as defined above.

Given its decision-making situation, an agent can compute its best rational action, a^*, as specified in Equation 1. This computation, however,

can be fairly complex, and certainly computationally expensive. In a multi-agent environment, for example, all of the information the agent has about the physical environment and about the other agents could be relevant and impact the expected utilities of alternatives courses of action. Sometimes the agent may have information about the other agents' state of knowledge, which is also potentially relevant (Castelfranchi and Muller 1993). Given these complexities, a mechanism for managing the agent's computational resources is needed. We suggest that emotional states provide for such ability, as we follow others in the field of AI who have started the tradition (Simon 1967; Sloman 1990; Picard 1997).

Assumptions of the Normative Rationalist Model

A variety of assumptions about decisions are made in the neoclassical rationalist model. We now describe some of the most salient ones, also discussed in a variety of works, in particular in Tversky and Kahneman (1986) and Johnson-Laird and Shafir (1993):

- An agent's choice between different options should depend only on those states in which they yield different outcomes. This is because if any state of the world yields the same outcome regardless of one's choice, it should be *canceled*, or eliminated from further consideration.
- Different but logically identical representations of the same choice situation, or different methods or eliciting a choice, should yield the same *invariant* preferences.

Another important assumption that the normative model makes in order to insure an agent's rationalist concerns consistency:

- An agent's *consistency* asserts that if an agent chooses action *a* today, he or she will choose *a* again tomorrow.

So far however, the model imposes no constraints on equilibrium. In order to get much an equilibrium, a further assumption has to be made at the macro-level, i.e., that all people are alike and that there does exist an objective probability distribution "out there." With this added assumption, preferences are assumed to be homogeneous among different individuals. That is to say that:

- An agents' *homogeneity* is established by assuming that individual agents, given the same information, will reach the same interpretation, will reason toward the same conclusion, will develop the same preferences and, finally, will act and react in a similar manner with the rest of the individual agents. Interestingly, as discussed in the next section, emotions are socially contagious and seem to account for such agents' homogeneity.

By assuming clone-like identical agents, with defined constraints and perfectly rational behavior, it becomes possible to globally predict individual behavior. From this global prediction, assessement of the state of the world can be made such that social equilibrium is reached. This is a theory referred to in agent modeling and economics as Rational Expectations.

- *Rational Expectations* are expectations formed by agents about other agents in a rational way, i.e., such that the value of the consequences of their choice is well-defined, established in advance, and shared by every other active agent. In this fashion, the expectations of individuals about a particular situation coincide in creating a stable world which validates them as predictions.[2]

  It is usually pointed out that resource limitations (e.g., time or energy), prevents maximization in the first place. In order to account for some quasi-rational behavior, variations of the original model came to life. These variations led to the theory refered to as *Bounded Rationality* (Newell and Simon 1972).
- *Bounded Rationality* takes into account the computational limitations of agents in assessing their *choice situation*, and addresses *the combinatorial explosion* of generating all the possible consequences of an action. Arthur (1994) describes the kind of strategies used to calculate the *value* of choices, such as inductive reasoning.

Advantages of the Normative Rationalist Model

While conscious that such assumptions about agents' preferences (and, therefore, actions) are somewhat unrealistic, the advocates of the neo classical model do not claim that such agents are real, but that *if the assumptions were valid* in the real world (or market), the neo classical model would give good predictions about the real world, with a high degree of certainty.

Another way the model is considered useful is to assume that the model is *not descriptive*, i.e., *not* about *how* people behave, *but rather prescriptive*, i.e., about how they *should* behave according to some rationalistic "etiquette" of normal behavior. This model of human behavior also posits that people will be very good at discovering ways in which they can advance their interests, even if it means being amoral, ignoring rules, breaking agreements, and employing guile, manipulation, and deception if they see personal gain in doing so.

The decision-theoretic rationalist model is the mathematical companion to the late eighteenth and early nineteenth century utilitarian philosophy, which imposes an "invisible hand" and equates rationality with self-interest, albeit self-damaging or irrational preferences might divert the agent from its actions.

Interestingly, even though it characterizes people as merely motivated by narrow selfish concerns and quite clever and unprincipled in their pursuit of their goals, theories based on perfect rationality are surprisingly successful in

generating explanations and specific predictions about organizations and economies.

Finally, the quantitative representations of probability and utility, along with the procedures for computing these, are also considered useful for the treatment of very simple decision problem, but they are less successful in more complicated and realistic cases found in everyday life.

View 1: Rationality Is the Opposite of Emotion

The very root of the normative rationalist model can be traced back to the major influence of seventeenth century French philosopher, René Descartes. Descartes, with the intention to question the dogmas of the church fathers by his method of systematic doubt, viewed the body as a complex automaton—a machine—and the mind as an independent entity with rules of its own and capable of generating thoughts and ideas. In 1637, his *Discourse on Method* was published, soon followed by his *Meditations*, both works strongly influencing the mind set of our entire Western culture:

> From that [truth "I think therefore I am"] I knew that I was a substance, the whole essence or nature of which is to think, and that for its existence there is no need of any place, nor does it depend on any material thing; so that this "me", that is to say, the soul by which I am what I am, is entirely distinct from the body, and is even more easy to know than is the latter, and even if body were not, the soul would not cease to be what it is. (*Descartes 1637*)

Historically, this split of the individual created what is typically referred to as the mind-body problem, but which Lisetti likes to refer to as the *anti-body problem*: the notion that people can exist without their bodies proper so that they can ignore (sometimes even disrespect) the close bidirectional interactions between body and mind. Cartesianism further led to the emergence of the domination of the rational scientific thought where "clear and distinct ideas" were valued more than the humanities which seemed to lack these "clear and distinct" qualities (Levi-Strauss 1978).

The legacy of the Cartesian identification of the mind with consciousness is that most people (unlike cognitive and social psychologists) believe that their conscious feelings and judgments control their actions. Instrospection "*obviously*" delivers them to us as part of immediate experience. Unfortunately, this claim is unstainable as most individuals are not aware of how they reason, nor sometimes of how they feel. Introspection and self-report have proven over and over to be unreliable in these matters (Mandler 1975). They can only provide cues to our underlying processes.

Furthermore, a definition of persons as thinking beings entails that emotion disrupts reason, and, therefore, if persons are to remain reasonable, that the influence of emotion must be removed from them, or they become romantic, helplessly led by their misleading and unruled emotions. Emotion, in this perspective, is understood to arise not from the mind, but from the body proper, a compelling force leading the persons away from the decisions they make, the reasons they have, the choices they make, and responsible for disrupting the rational calculations they could perform, if only emotions did not get in the way.

To be sure, affective phenomena such as moods and emotions can often influence us negatively when these are experienced as "out of control" or turn into disorder (Loewenstein 1996). During the Dark Ages of the Cartesian rationalist tradition, this dysfunctional view of emotions has been the sole one considered.

It is currently fully acknowledged scientifically (Clark 1996), however, that Cartesianism created an unnatural split between the mind and the body, between the senses and the intellect, and between reason, rationality, and emotions. In his famous, "*Je pense, donc je suis*," Descartes placed the human reason at the center of human being, not only discarding the emotional realm of life as a central part of existence, but also mindlessly ignoring the nature of the "I" that thinks.

Our insistence on the significance of emotion in social and cognitive processes is not necessarily, however, an acceptance of a romantic disposition which rejects the assumption that society and individual could be ordered by rational principles and instead acclaims emotion as the sole basis of value and conduct. This is the reason we feel the need to introduce a new notation or term, namely, $rational_1$, to define a way in which people can be both rational, i.e., acting reasonably, and emotional, i.e., taking advantage of the affective phenomena they experience during rational problem-solving situations.

WHAT *IS* A RATIONAL AGENT? TOWARD DESCRIPTIVE $RATIONAL_1$ AGENT MODELING

We are motivated by the desire to depart from the durable three-century-old Cartesian rationalist tradition, which holds (among other tenets) that emotion is the opposite of rationality. Such view is ultimately unsustainable because those who wish to suppress emotion in fully realizing reason, are typically engaged by an emotional commitment to the project.

In the conventions that shape our Western thoughts, rationality and emotion are alternatives—one is defined by what the other is not. However, there are at least two other possible relationships between emotion and rationality, which are much more creditable than the one usually claimed to be. These two views are that (i) rationality requires emotional guidance, and

(ii) that rationality and emotions are continuous. But, first, let us introduce a new notation that will make our perspective easier to grasp.

Humans Are $Rational_1$

We continue the debate about what it means to be rational in human reasoning and decision making by acknowledging, like Evans et al. (1993) before us, the confusion that two implicit definitions of rationality has brought about.

We suggest to rename "rationality" as $rationality_1$, such that it refers to reasoning in a way which helps one to achieve one's goals as opposed to $rationality_2$, which refers to reasoning in a way which conforms to a supposedly appropriate normative system such as formal logic, for example (Evans et al. 1993).

Our new term $rationality_1$ can now refer to phenomena not considered previously in the definition of rationality (usually understood as $rationality_2$ above). In particular, $rationality_1$ includes the role that emotions play in rationality in their plethora of ways during the human decision making process.[3]

In Lisetti and Gmytrasiewicz (2000), we adumbrated an approach toward a more descriptive model of rational agents by including the role of emotions in their decision-making situations and actions. We introduced the notion of *emotional transformation* which we now describe, and which we will connect to our views on the role of emotions in $rationality_1$ in the next sub-sections.

We assume that the emotional transformations themselves are triggered by some input, *IN*, that the agent experiences, which can be external (e.g., the sight of a face) or internal (e.g., the realization that an emotion is justified).

Let us denote as **D** the set of all decision situations *D*, as defined previously; we postulate that each $D \in \mathbf{D}$ corresponds to an emotional transformation. Further, let **IN** be the set of all possible inputs.

Emotional transformation definition. An *emotional transformation* is a function $EmotTrans: \mathbf{D} \times \mathbf{IN}^* \rightarrow \mathbf{D}$.

An *emotional transformation*, therefore, associates an emotional state with a decision-making situation. An emotional transformation changes one decision-making situation, where an agent is in an emotional state, into another one. Affect (less fine-grained than emotion, explained in the next section) is experienced primarily and ongoingly (Zajonc 1984).

Given an initial emotional state, *D*, and a (possibly empty) history of inputs *IN*, the value of the *EmotTrans* function is $EmotTrans\ (D, IN) = D'$, where D' is the agent's new emotional state. D' may differ from *D* in a number of ways. We look at some possibilities that correspond to some of the more intuitive emotional states in the next two sub-sections.

It is important to note that the association of some emotions with such transformations does not account for all of the richness of emotions in humans, but rather concerns only emotions that impact the agent's decision-making processes. Aesthetic emotions, such as *awe*, for example, are not considered in this framework.

We now discuss the two new approaches to the relationship between emotion and rationality: (i) emotion supports rationality by providing it with salience and goal formation, and (ii) emotion and rationality are continuous with each other, and offer different ways of looking at the same thing.

View 2: *Rationality*$_1$ Requires Emotional Guidance

Indeed, contrary to the Cartesian rationalist tradition, *rationality*$_1$ and emotion can be considered not as necessarily opposed, but clearly different faculties, and their differences can be considered as allowing each to serve a division of labor in which their distinct capacities contribute to a unified outcome.

In our current article we develop a formal framework within the original rational agent paradigm which, fully acknowledging that the distinction between *rationality*$_1$ and emotion is at least blurred and not sharp, formalizes some of the interactions between emotion and *rationality*$_1$ in the decision-making process illustrated in Figure 2.

1. Emotions Influence Attention Processes: Transformations of the Set of the States of the World S

There is growing evidence in neuroscience that there is no "pure reason" in the healthy human brain—emotions are vital for healthy rational human thinking, behavior, and decision making (Damasio 1994). There are descending and ascending pathways in the neural mechanisms of emotions which intertwine with other mechanims responsible for cognitive processes, such that they cannot really be considered separately (Derryberry and Tucker 1992). While the descending pathways influence processes, such as attribution and appraisal, the ascending pathways are strongly linked with memory and attention. We illustrate this notion in Figure 2 with our dotted arrow (labeled 1), from "Affect" to "Sensors/Percepts."

Formally, these are transformations $EmotTrans(D, IN) = D'$ such that:

$$D = \langle P_c(\boldsymbol{S}), A, Proj, U\rangle, \quad \text{and} \quad D' = \langle P_c(\mathbf{S}'), A, Proj, U\rangle.$$

An emotional transformation that implements the perceptual reduction of the possible states of the world is one in which $S' \subset S$. For example, the effects of anxiety on attention (and the effects of depression on recall for that matter) have been studied extensively (Williams et al. 1996). Anxiety can

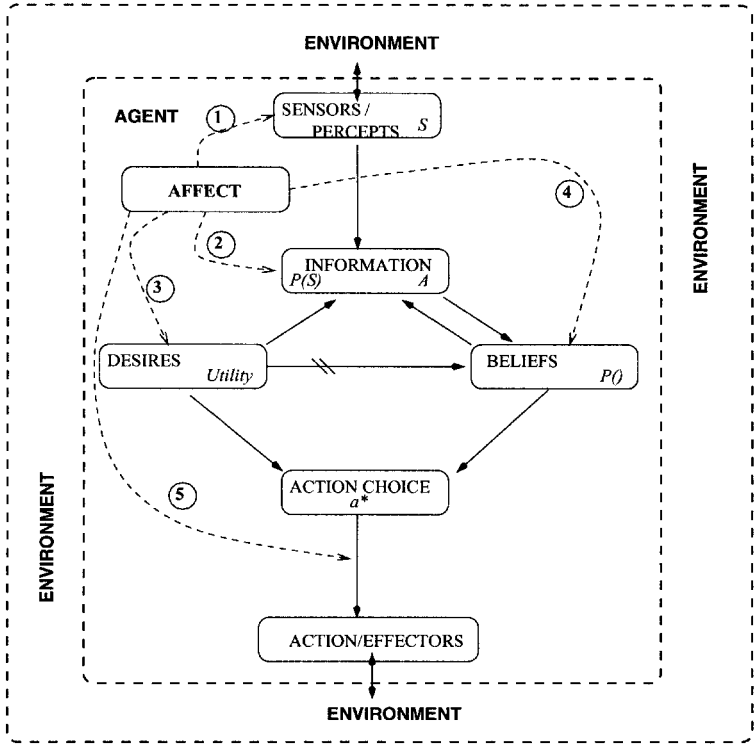

FIGURE 2. *Rational*$_1$ utility-based agent with affect.

increase the chances of perceiving negative information, whereas depression has been found to lead to recall of more negative information. Therefore, anxiety and depression bias information processing towards negative aspects of the self and of the environment (Daniels and Guppy 1997). When an agent passes from an emotional state of say, sadness, to one of anxiety, the agent reduces the set of the states of the world to $S' \subset S$, where S' contains only states associated with mostly negative valence.

Interestingly, the same is true about positive states. Enthusiasm is associated with optimistic evaluations of the environment (Munz et al. 1996) and greater creativity (Estrada et al. 1996). So when an agent passes from an emotional state of say, neutral, to one of joy or enthusiasm, the agent reduces the set of the states of the world to $S' \subset S$, where S' contains mostly states associated with positive valence.

2a. Emotions Enable Us to Choose Among Options None of which Is Rationally Superior to the Other: Transformations of the Action Space A.

de Sousa (1990) summarizes this view by writing that "Emotions are among mechanisms that control the crucial factor of *salience* among what

would otherwise be an unmanageable plethora of objects of attention, interpretation, and strategies of inference and conduct.'' Information and action, therefore, cannot organize themselves, and a crucial organizational function is performed by emotions. This notion is represented in Figure 2 by the dotted arrow (labeled 2) from ''Affect'' to ''Information.''

The *somatic markers* hypothesis proposed by Damasio (1994) brings another contribution to the notion that emotional guidance helps rationality. *Somatic markers* (those emotionally borne physical sensations which ''tell'' those who experience them that an event is likely to lead to pleasure or pain) precede thought and reason. They do not replace inference or calculation, but they enhance decision making by drastically *reducing the number of options* for consideration.

> At their best, feelings point us in the proper direction, take us to the appropriate place in a decision-making space, where we may put the instrument of logic to good use. (*Damasio 1994*)

A transformation of the action space A is achieved, for example, by narrowing the set of alternative actions considered to encompass only a small subset of all of the actions, predisposes the agent to take action from this smaller set. This constitutes the *action tendency* that the emotion invokes in the agent, as postulated by Frijda (1986), and explained in further detail in the section ''Functional Attributes and Action Tendencies.''

Formally, these are transformations $EmotTrans(D, IN) = D'$ such that:

$$D = \langle P_c(S), \mathbf{A}, Proj, U \rangle, \quad \text{and} \quad D' = \langle P_c(S), \mathbf{A}', Proj, U \rangle.$$

An emotional transformation that implements an action tendency is one for which $A' \subset A$. For example, an agent becoming angry may result in it considering only a subset of its behavioral alternatives, say, ones of aggressive nature. In the extreme, narrowing the set A to a single action implements a behavioral condition-response rule, similar to very basic reflex-like affective state (e.g., avoid pain, seek pleasure). Such emotional transformation is obtained when A' is a singleton set, containing only one behavior. This is an implementation of an emotional condition-action rule, which results in the agent's being capable of performing only a single behavior.

Another intuitive special case of such transformation is one that results in the agent's deliberating in a more short-term fashion, such as if being rushed or panicked under time pressure. Formally we have: $\forall a_i' \in A' : t_{a_i'} \leq t_{a_i}$, which states that the time horizon of alternative plans considered has diminished. This is characteristic of human decision makers; people frequently become more shortsighted when they are rushed or panicked, since they have no time to consider long-term effects of their alternative behaviors.

2b. Emotions as Tie-Breaker Can Also Be Formalized with Transformations of the Probabilities of States

The idea behind this transformation is that changing these probabilities, for instance, by simplifying them, can be helpful, and save time under time pressure. Formally, these are transformations $EmotTrans(D, IN) = D'$ such that

$$D = \langle \mathbf{P_c}(\mathrm{S}), A, \mathbf{Proj}, U \rangle, \quad D' = \langle \mathbf{P'_c}(\mathrm{S}), A, \mathbf{Proj'}, U \rangle.$$

The most radical simplification is one that makes the most likely state to be the only possible state or result. This corresponds to considering only the most likely result of action and neglecting all less likely states and is often observed in human decision makers.

3. People Act Rationally to Acquire Optimal Emotional Dispositions and/or to Maximize Their Positive Emotional Experiences: Transformations of the Utility Functions U

The intuition behind this transformation is that emotions and feelings both implement U, as well as modify it. Humans evaluate desirability of states by having positive or negative feelings about them. This notion is represented in Figure 2 by the dotted arrow (labeled 3) from "Affect" to "Desires."

Positive or negative emotions, or moods, may alter these evaluations by, say, decreasing them, as in melancholic or depressed moods (when everything looks bleak), or increasing them, as in elated or happy moods. It has indeed been observed that when people are happy, their perception is biased at selecting happy events, similarly for negative emotions (Bower 1981).[4]

Frequently, it is convenient to represent the utility function, U, as depending on a small number of attributes of the states of the world, as opposed to depending on the states themselves. For example, some humans may prefer, say, all of the states in which they have more money, and are more famous. The attributes, say wealth and fame, are then convenient factors in terms of which the utility function can be expressed.

The formalism can also be used to model agents who value inner state quality (versus external material quantity), such as feeling healthy, loved, and loving. In this case, the attributes are then subjective attributes based on the actual states of feeling loved and healthy. Multi-attribute utility theory postulates that, in some simple cases, the utility of a state is a weighted sum of the utilities, $U(X_l(s))$ of individual attributes:

$$U(s) = \sum_{X_l \in Attributes} W_{X_l} U(X_l(s)), \tag{2}$$

where the W_{X_l} is the weight, or intuitively, the importance or desirability of the attribute X_l. Having the weights of the attributes explicitly represented is convenient since it enables the tradeoffs among the attributes the agent may have to make. For example, the agent may have to give up some of its health to improve its fame.

Emotional states can change the weights of the factors contributing to the utility ratings in Equation 2 above.

Formally, these are transformations $EmotTrans(D, IN) = D'$ such that

$$D = \langle P_c(S), A, Proj, \mathbf{U} \rangle, \quad \text{and} \quad D' = \langle P_c(S), A, Proj, \mathbf{U}' \rangle.$$

A melancholic mood, for example, can result in the evaluation of the desirability of every state to diminish: $\forall_s \in S : U'(s) \leq U(s)$.

In severe depression, furthermore, no state of the world is desirable: $\forall_s \in S : U'(s) = 0$.

Mellers et al. (1997) studied how unexpected outcomes have greater emotional impact than expected outcomes and decision affect theory includes utilities, expectations, and conterfactual comparisons into hedonic responses. Choices between risky options are considered as maximization of expected emotional experiences such as regret, disappointment, and surprise.

Lastly, an interesting study of Buddhism as a theory of how to maximize utility (i.e., minimize suffering in Buddhist terms) under constraints that are internal to the agent (self-control), rather than external, is presented in Kolm (1986). For instance, one can calculate the optimal allocation of time among various activities, such as meditation (character modification via desire inhibition or emotion dissociation), working in order to consume, and consumption itself.

View 3: $Rationality_1$ and Emotions Are Continuous

Lastly, emotions can permit action that would be inhibited if it were to rely on logic or calculation alone.

4. Emotions Provide Us with Gut Feelings which Help Us Form Rational Beliefs

Choice, as we have seen, can be based upon beliefs. Beliefs can be formed rationally by applying strick rule reasoning. More often than not, however, beliefs are formed from an interaction of high reason with emotional states. This notion is represented in Figure 2 by the dotted arrow (labeled 4) from "Affect" to "Beliefs."

More importantly, many pieces of information that we possess are not consciously acknowledged (contrary to the Cartesian assumption that introspection makes it all available as discussed earlier in this paper. Furthermore, cognitive basis of the emotions include unconscious knowledge

(Zajonc 1980). That is to say that emotional reactions are cues to our unconscious assessment of a situation (Rumelhart and Lisetti 1996).

Johnson-Laird and Shafir (1993) and Johnson-Laird and Oatley (1992) have reminded the cognition community of the inability of logic to determine which of an infinite number of possible conclusions are sensible to draw given a set of premises.

Indeed, how do people decide which path to take given some evidence? There is not always time to consider *every* possible logical constraint and the path associated with it. Emotions do not merely provide a tie-breaker in making decisions. Rather, emotions play an essential role in *learning the biases* required to construct rational reponses.

Inappropriate, ambiguous, and competing goals, on one hand, and imperfect, disorganized, or absent knowledge, on the other hand, undermine reason and rationality, whereas emotion may clarify or *define goals* and "bridge" information. This latter is possible because inherent in a goal is an emotion.

For example, take when one meets someone who makes one feel vaguely uncomfortable. One cannot necessarily formulate the belief about the person that justifies that emotion, but one can infer from that emotion that one has such a belief. That belief can then serve as a premise for action as shown in Figure 2, for example, the action not to see that person again.

5. Emotions Enable Us to Take Action once the Action Has been Chosen

From James' 1897 account, emotions contribute to rationality by providing a feeling of certainty concerning the future, which is necessary if action is to occur and the agent to proceed to implement its chosen action.[5]

This notion is represented in Figure 2 by the dotted arrow (labeled 5) from "Affect" to the arrow connecting "Action Choice" to the action itself drawn as "Action/Effectors."

People sometimes achieve a goal in part because they do not try to maximize utility in the technical sense, or hold coherent and consistent beliefs. With bounded capacities and limited time, people would sometimes miss the chance to achieve a reasonably satisfying goal if they paused to wonder whether their preference relation or their beliefs satisfied some normative principles. They can sometimes save precious time by accepting a plausible conclusion without examining closely the logic of the argument (Evans et al. 1993), or by being very vague about their preferences (Zajonc 1980).

6. Emotions Are Socially Contagious

From a purely economic perspective, the market is often referred to by traders (as opposed to academic theorists) as possessing "moods," sometimes describing it as "nervous," "sluggish," or "jittery" (Arthur 1995).

These macro-economic moods are also a product of individuals' personal emotional states. If investors feel *anxious* about the market, they may behave differently than if they are *relaxed* and *confident*. Emotions felt by individuals such as *anxiety, surprise, excitement*, or *satisfaction*, and even the more pathological ones, such as *hysteria, hypocondria* or *panic*, are often at the root of sudden economic market shifts.

Primitive emotional contagion, in which one person's affective state influences another's regardless of the situation, has also been studied extensively and can provide insights as to how to model distributed artificial *rational*$_1$ agents (Gurtman et al. 1990; George 1990; Hatfield et al. 1986).

7. *Emotions Are Socially* Rational$_1$

The importance of the articulation of appropriate goals or purposes for rational action to occur, and the contribution of emotion to this process, are discussed in Hirshleifer (1987) and Frank (1988). Frank explains how self-interest theory is flawed in its inability to discriminate between appropriate and inappropriate means which might be used to satisfy the agent's interests. Evidence is provided that social morality, even though it may appear irrational, nevertheless confers real advantages in the long run.

An agent's emotional *commitment* leads the individual from a narrow concept of self-interest to a broader concept of self-interest, but the notion of self-interest itself remains. From this perspective, simply unselfish, non-opportunistic, even altruistic actions in the end will yield greater material benefits to the agent. Emotions provide appropriate or enhanced goals for self-interested action, enlarging its rationality.

It has been observed that only narrowly self-interested utilitarian agents maximize expected utility. Concerns about the variety of unethical practices and actions that would lead to a maximal return (e.g., undetected cheating) can motivate people to choose less than optimal options in terms of utility function calculations, and by definition act irrationally—cheating confers advantage, and to avoid it, therefore, avoids a means of satisfying self-interest.

Here the ability to role-play the consequences of an action and to anticipate the shame felt when one is caught cheating, for example, can explain how people might act irrationally according to the traditional model, but *rationally*$_1$ in ours. A good account of this approach is also discussed in Barbalet, (1998), who shows how emotions are central to social order and conformity, rationality, human rights, conflicts of social inequalities, the processes of social action and structural change. Other contributions point at the role of emotions (current and anticipated ones) in the decision-making process (Tsiros 1998; Connolly et al. 1977; Zeelenberg et al. 1998; Bohm and Pfister 1996; Burnett and Lunsford 1994; Pfister and Bohm 1992).

Now that we have described the various transformations that emotions can bring to a decision-making situation, we turn to knowledge representation in order to formulate a scheme to represent the affective information and processes needed for such transformations to occur.

AFFECTIVE KNOWLEDGE REPRESENTATION (AKR) AND DYNAMICS FOR $RATIONAL_1$ AGENT MODELING

We created AKR to enable the design of a variety of artificial, autonomous (i.e., self-motivated), socially competent agents, so they can measure and predict the emotional state of other agents, as well as be guided by their own, should they need to make decisions. Socially intelligent autonomous agents are applicable to a variety of fields, from human-computer interaction and user-modeling (Castelfranchi et al. 1997; Hayes-Roth et al. 1998; Lisetti et al. 1998; Lisetti and Schiano 2000; Paiva 2000; Picard et al. 2001; de Rosis 2001; Lisetti and Bianchi-Berthouze 2002; Lisetti [in press]; Hudlicka and McNeese 2002), to education (Conati 2002), to entertainment (El-Nasr et al. 1999), to telemedicine (Lisetti et al. 2001), to multi-agent systems (Gmytrasiewicz and Durfee 2000) and distributed AI, and to robotics (Sloman and Croucher 1981; Lisetti 2002; Velasquez 1999; Breazeal 1998; Michaud et al. 2000; Murphy et al. [in press]).

Affect Taxonomy

We combined and reconciled aspects of the main current theories of affect, mood, and emotion (Zajonc 1984; Ortony and Turner 1990; Ekman 1993; Frijda 1986; Wierzbicka 1992) into a simplified comprehensive, (but not complete) taxonomy of affect, mood, and emotion sketched out earlier (Lisetti 2002). Our taxonomy of affective states in Figure 3 is aimed at differentiating among the variety of affective states by using values of well-defined componential attributes.

Personality. Personality represents characteristics of an autonomous (i.e., self-motivated) organism that account for consistently chosen patterns of mental reaction including behavior, emotions, and thoughts over situations and time (Moffat 1997).

As shown in Figure 3, because emotions are at the bottom of the hierarchical model, emotions do not necessarily imply specific personalities, since some emotions might be experienced by different types of agents (artificial or natural). The type of personality is inherited from the higher node in the tree, allowing agents of different personality type to still experience the full range of possible emotions, as advocated by other computational approaches (Castelfranchi 1997).

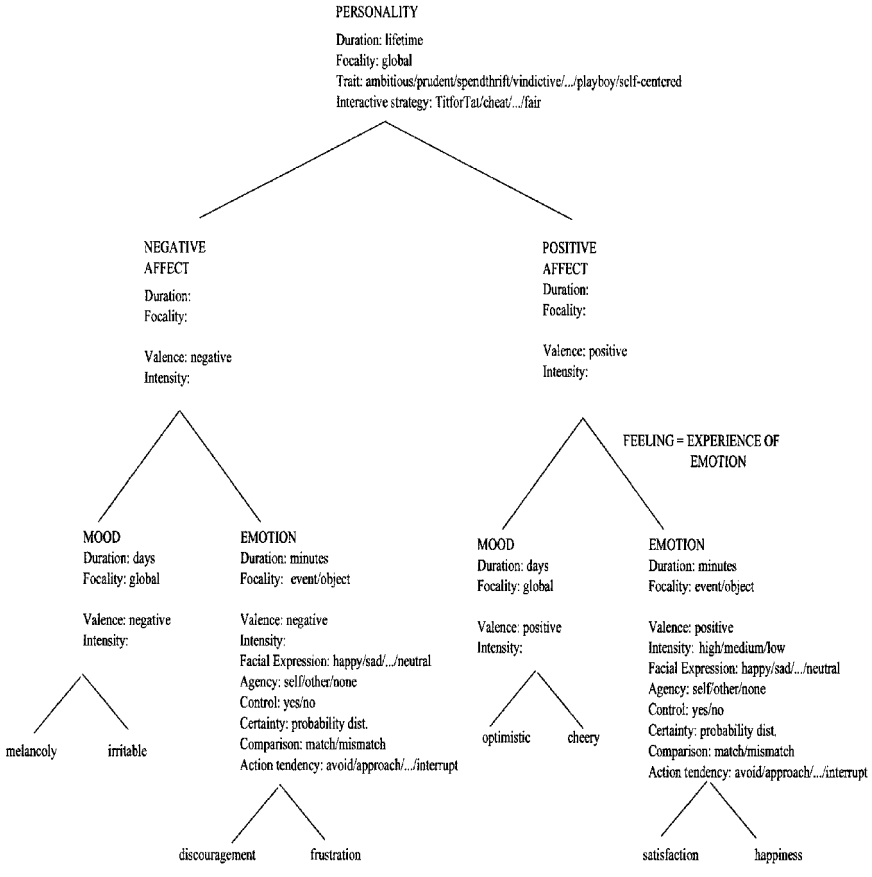

FIGURE 3. Hierarchical model of personality, affect, mood, and emotion.

Furthermore, because we adopt the functional view of emotions, which identifies emotions as related to goals and action tendencies (such as self-preservation, avoid negative experiences, approach enjoyable things, etc.), our model is compatible with goal-oriented theory of personality (Carbonnell 1980).

In addition, because the interactive strategies (tit-for-tat, cheat, etc.) are specified in the model at a higher level than at the emotion level, and because personality explicitly is represented as lasting a lifetime and not related with any specific event, this approach is in agreement with other views which emphasize the main distinction between personality (stable) and emotion (changeable) (Castelfranchi 2000). Our approach, however, does not preclude the notion that different personalities can influence an agent's propensity toward a particular set of emotions and moods.

Affect. Affect varies along two dimensions: *valence*, which can be positive or negative (the pleasant and unpleasant dimension), and *intensity*, which varies in terms of degree.

Mood. Moods are affective phenomena encoding coarser-grained information, of longer duration than emotions (days), and without a specific focus nor cause. Drug intake or hormonal level changes can alone cause specific moods.

Emotion. Emotions are multimodal, physiological, and mental fine-grained affective phenomena which are *about* something (e.g., a person or an event). Emotions are changes in activation of behavioral dispositions or transformation of dispositions to act, caused by relevant events or circumstances.

Each emotion is considered as a collection of emotion components, such as its valence (the pleasant or unpleasant dimension) or its intensity (mild, high, extreme), etc. Some of these can be measured via pattern recognition techniques when referring to a human user, others can be computed by combining the observed emotional signals and stored information known about emotion dynamics encapsulated in the model.

In our representation, we also included the action tendency of each emotion (Frijda 1986), which corresponds to the signal that the emotional state experienced points to: a small and distinctive suite of action plans that has been (evolutionarily) selected as appropriate, e.g., approach, avoid, reject, continue, change strategy, etc.

Feeling. Feelings are the mechanisms that enable an emotion to be experienced subjectively, bringing awareness to the agent's inner states such that it can be conscious of them, and possibly volunteer to express them to others.

Emotion Components

In order to address some of the difficulties of the previous computational approaches to emotion modeling pointed out by Pfeifer (1988), namely the lack of representation of physiological and subjective parameters, we do not split "emotion" and "cognition," but rather merge them into a structure that encapsulates simultaneously each of the three phenomena accompanying emotions:

1. Autonomic Nervous System (ANS) arousal, signaling affect intensity, and valence.
2. Expression, for now only facial expression is included because our user model uses facial information pattern recognition techniques (Lisetti and

Schiano 2000; Lisetti and Bianchi-Berthouze 2002; Lisetti [in press]). However, it could also include vocal and body posture.
3. Subjective experience, including cognitive appraisals (such as modal beliefs, criteria, standards, etc.).

In an effort to identify what makes one emotion different from another, we include elements from the "cognitive view" of emotions, which advocates a componential approach (Leventhal and Scherer 1987; Frijda and Swagerman 1987; Frijda 1986; Ortony et al. 1988; Roseman et al. 1996). From this approach, cognitive structures associated with emotions are considered to represent the subject's checks—appraisal or evaluation (Castelfranchi 2000)—of the events confronting them.

These checks are part of the subjective experience of emotion and can be represented with a limited number of components. Each type of checks is described as a unique pattern of such components, or dimension values. As with the set of basic emotions, which varies among theories, several dimensions are often considered, but the following are found in most analyses: valence (pleasantness or unpleasantness), intensity/urgency, agency/responsibility, novelty, controllability, modifiability, certainty, external social norms compatibility, and internal standards compatibility. We also included duration and focality which differentiate emotions from moods, for future potential expansion modeling moods as well.

These components are listed below:

- **Facial expression** (*happy/sad/surprised/disgusted/angry/fearful/neutral*). Used to store the facial expression associated with the emotion. Some emotions are not associated with any specific facial expression (neutral), or can vary among cultures and individuals.
- **Valence** (*positive/negative*). Used to describe the pleasant or unpleasant dimension of an affective state. Each affective phenomena is associated with a valence, except for the emotion of surprise, which can be either positive or negative depending upon the nature of the stimulus.
- **Intensity** (*very high/high/medium/low/very low/null*). Varies in terms of degree. The intensity of an affective state is relevant to the importance, relevance, and urgency of the message that the state carries.[6]
- **Duration** (*lifetime/days/minutes*). Used to indicate that moods are more muted and last longer than emotions, which is indicated by the *duration* attribute measured in terms of days, as opposed to minutes in the case of emotions; it can also be used to resolve the conflict between personality and emotion by assuming that the underlying mechanisms are essentially the same, and that only the time-course and temporal of their influence varies; personalities can be permanent and last a lifetime.

- **Focality** (*global/event/object*). Used to indicate whether the affective phenomena is *global* (the cause may not be a meaningful event, but rather a biochemical change) as in moods in which the cause has become detached from the felt action readiness (the cause may not be an experienced meaningful event, but it may be biochemical); or, on the other hand, as in emotions which are mostly about something: an *event* (the trigger to surprise) or an *object* (the object of jealousy). Globality can also differentiate emotions: depression from sadness, bliss from joy, anxiety from fear. In depression the world as a whole appears devoid of intentional objects; similarly in happiness, the environment as a whole appears tinted with positive valence.
- **Agency** (*self/other/nature*). Used to indicate who was responsible for the emotion, the agent itself *self*, or someone else *other*. For example, if the agent is angry at someone, the agent parameter will point to that person; but if the agent is depressed, agency will most likely point to self.
- **Novelty** (*match/mismatch*). Used to refer to whether a novel and unexpected stimulus occurred causing mismatch with the subject's expectations regarding the stimulus triggered.
- **Intentionality** (*other/self/unspecified*). Used to refer to whether the triggering event is perceived as caused by some live intending agent. In anger, it is *other*, whereas in self-hatred and guilt, it is *self*.
- **Controllability** (*probability distribution*). Used to refer to how much the agent believes she, he, or it can control the current situation. Controllability is the component that turns danger from threat into challenges and, therefore, negative into positive emotion. Change from angry protest to despair and resignation can be interpreted as a consequence of the fact that uncontrollability gradually drawns. Daniels and Guppy (1997) shows that people with negative emotional symptoms tend to shift to a more external locus of control, such as feeling that their fate is controlled by external circumstances rather than by themselves. This is represented by the controllability component set to *none*.
- **Modifiability** (*probability distribution*). Used to refer to duration and time perspective, or to the judgment that a course of events is capable of changing. Modifiability carries with it the past, in the sense that what has been for a long time may well be forever. It can also apply to current events, e.g., suffering a situation as if it will never end or a feeling of self-confidence.
- **Certainty** (*certain/uncertain/non-uncertain*). Used to refer to anticipation of effects to come, and how (subjectively) certain the subject is about the consequences of the situation. For example, joy implies absence of uncertainty (uncertainty about how friends will respond takes away the joy of going to meet them), yet the aspect of certainty is implicit (hence our three values).

- **Legitimacy** (*yes, no*). Used to indicate whether the emotion is experienced as a legitimate state.
- **External (social) norm** (*compatible/uncompatible*). Used to refer to whether the event (usually an action) conforms to social norms, cultural conventions, or expectations of significant other.
- **Internal (self) standard** (*compatible/uncompatible*). Used to refer to whether the event (usually an action) is consistent with internalized personal standards as part of the self concept or ideal self.
- **Action tendency**. Identifies the most appropriate (suite of) actions to be taken from that emotional state.
- **Causal chain**. Identifies the causation of a stimulus event (described next).

The summary of the list of components associated with their possible values is shown in Table 1.

Functional Attributes and Action Tendencies

From the Darwinian categorical theory of emotions (Darwin 1872), emotions can be discretely categorized. Emotions are considered as mental and physiological processes, caused by the perception of general categories of event, that elicits internal and external signals and a matching suite of action plans. This Darwinian perspective proposes that bridging the gaps of rationality becomes possible if many specific emotional states are mapped into a few broad classes of reaction, or *action tendencies*.

Action tendency. Emotions which are called ''primary'' or ''basic'' are such in the sense that they are considered to correspond to distinct and elementary forms of action tendency. Each ''discrete emotion'' calls into readiness a small and distinctive suite of action plans that has been selected as appropriate when in the current emotional state. Thus, in broadly defined recurring circumstances that are relevant to goals, each emotion prompts both the individual and the group in a way that has been evolutionarily more successful than alternative kinds of prompting.

The number and choice of what is called basic or primary emotions vary among various emotion theories, and we have selected the ones that seem to reoccur consistently across emotion theories. Their associated action tendency are listed in the Table 2.

The emotional signal that is sent when a sub-goal is achieved acts to prompt the individual to continue with the current direction of action, whereas the signal sent when a goal is lost, indicates a need to change the course of action, or to disengage from the goal. Ensuing actions can be communicated to others in the same social group, which in turn, can have emotional consequences for these other individuals too.

TABLE 1 Each Component has a set of Possible Component Values. Some Component Values are Expressed as a Range of Scalars for Clarity Sake, but they Actually Correspond to Probability Distributions. The Default Value for Duration is set to *minutes* for Emotion, for Certainty to *non-uncertain*, and for the other Components it is *Unspec*.

Pre-defined emotional categories Emotion components:	Component values:
Emotion label:	Frustration, relief, disappointment, surprise, amazement, fear, anger, indignation, shock, hurt, remorse, guilt, shame, humiliation, embarrassment, pride, sadness, distress, sorrow, grief, despair, joy, contentment, excitement, satisfaction, happiness, interest, disgust, indifference/boredom
Facial expression:	Happy, sad, surprised, disgusted, fearful, angry, neutral, unspec.
Valence:	Positive, negative, unspec.
Intensity/urgency:	Very high, high, medium, low, very low, none, unspec.
Duration:	Minutes (default), days, lifetime
Focality:	Event, object, global, unspec.
Agency:	Self, other, nature, unspec.
Novelty:	Match, mismatch, unspec.
Controllability:	High, medium, low, none, unspec.
Modifiability:	High, medium, low, none, unspec.
Certainty:	Certain, uncertain, non-uncertain (default)
External norm:	Compatible, incompatible, unspec.
Internal standard:	Compatible, incompatible, unspec.
Action tendency:	Change strategy, reorient/interrupt, avoid, approach, remove obstacle, free activate, inactivate, excite, attend, reject, nonattend, retain control, regain control, submit, prepare, chunkdown

Emotion Beliefs and Causal Chains

We adapted the semantic meta-definitions of emotion concepts developed by Wierzbicka using language independent primitives (Wierzbicka 1992) to create the *causal chain*.

Causal chain. A causal chain of events describes the *subjective cognitive experience* components, which are associated with the emotion, the beliefs, the goals, and their achievement or lack of. These in turn can also be formalized as shown by Castelfranchi (2000). Some examples are shown at the bottom of Table 3.

Putting It All Together into Probabilistic Frames

We use the formal probabilistic frame-based representation scheme described in Koller and Pfeffer (1998), which integrates the classical frame-representation systems (Minsky 1975) (limited because of their inability to deal with uncertainty), and Bayesian networks (limited in their ability to handle complex structured domains). This approach preserves the advantages of both schemes and adds to them by allowing uncertainty over the set

TABLE 2 Action Tendency Table

Action tendency	End state	Function	Emotion
Approach	Access	Permitting consummatory activity	Desire
Avoid	Own inaccessibility	Protection	Fear
Attend	Identification	Orientation	Interest
Reject	Removal of object	Protection	Disgust
Agnostic	Removal of obstruction	Regain of control	Anger
Interrupt	Reorientation	Reorientation	Shock, surprise
Free activate	Action tendency's end state	Generalized readiness	Joy
Inactivity	Action tendency's end state	Recuperation	Contentment
Inhibit/prepare	Absence of response	Caution	Anxiety

of entities present in the model, and uncertainty about the relationships between these entities.

Each emotion is encapsulated into a *frame*, each emotion feature or component is allocated a *slot*, each of which may have *slot values* (or *fillers*). A slot represents a binary relation on frames: if the filler of slot A in frame X is Y, then the relation A(X,Y) holds. Slots can be multi-valued or single-valued. Each slot in a frame may have associated *facets*. A facet is a ternary relation: if the facet value F on slot A in frame X is Y, then the relation F(X, A, Y) holds. A standard facet is value-type, which specifies a value restriction on the values of a slot.

The advantage of this representation, as shown in Table 3, is that it allows for all the components of emotion to be represented and associated with probabilistic values given the uncertain nature of some of these. For example, an agent might be fairly certain about the agent responsible for the frustration: *agency* is associated with *other* by a 1.0 probability of certainty. The agent, however, could be fairly uncertain about whether it has control over the current situation, expressed with the probability distribution for *certainty* shown in Table 3.

A Markov Model of Emotional States Dynamics

Now that we have described how each emotional state is associated with a set of features or components, let us turn to the dynamics between these various states. A dynamic model of the possible transitions between various states can not only assist our system in identifying the agent's current state when, say, pattern recognition processes fail to recognize it, but it can also assist in predicting an agent's most probable future emotional state given its current state.

TABLE 3 Probabilistic Frame for Frustration

Frustration
Emotion components:
Simple slot: emotion label = *frustration*
Facet: type string
Multivalued complex slot: facial expression
Facet: type {happy, dad, surprised, disgusted, fearful, angry, neutral, unspecified}
Facet: distribution {0.0, 0.2, 0.0, 0.0, 0.0, 0.6, 0.1, 0.1}
Simple slot: valence = *negative*
Facet: type positive, negative, unspecified
Multivalued slot: intensity
Facet: type {very high, high, medium, low, very low, none, unspecified}
Facet distribution {0.15, 0.15, 0.15, 0.15, 0.15, 0.1, 0.15}
Simple slot: duration = *minutes*
Facet: type {minutes, days, lifetime}
Multivalued complex slot focality
Facet: type{event, object, global, unspecified}
Facet: distribution {0.1, 0.9, 0.0, 0.0}
Multivalued complex slot: agency
Facet: type {self, other, nature, unspecified}
Facet: distribution {0.0, 1, 0.0, 0.0}
Multivalued complex slot: novelty
Facet: type {match, mismatch, unspecified}
Facet: distribution {0.45, 0.45, 0.1}
Multivalued complex slot: controllability
Facet: type {high, medium, low, none, unspecified}
Facet distribution {0.1, 0.1, 0.3, 0.4, 0.0}
Multivalued complex slot: modifiability
Facet: type {high, medium, low, none, unspecified}
Facet: distribution {0.1, 0.1, 0.4, 03, 0.1}
Multivalue complex slot certainty
Facet: type {certain, uncertain, non-uncertain}
Facet: distribution {0.4, 0.3, 0.3}
Simple slot external norm = *unspecified*
Facet: type {compatible, incompatible, unspecified}
Multivalue slot: Internal Standard
Facet: type {compatible, incompatible, unspecified}
Facet: distribution {0.6, 0.25, 0.15}
Multivalued complex slot: action tendency
Facet: type {change strategy, reorient/interrupt, . . . ,avoid}
Facet: distribution {1,0, 0.0, . . . ,0.0}
Causal Chain:
I want to do something
I can't do this
because of this I feel bad

External Event as Inputs

We use Markov Models to represent various emotional states dynamics. Picard (1997) introduced this notion earlier. As shown in Figures 4 and 5, we can represent with a Markov Model a process that goes through a series of states. The model describes all the possible paths through the state space and assigns a probability to each one. The probability of transitioning from the current state to another one depends only on the current state, not on any prior part of the path.

A Hidden Markov Model (HMM) is even more useful in our case because it is just like a Markov Model except that each state has a probability distribution of possible outputs, and the same output can appear in more than one state.

We are adapting the idea to various specific contextual situations and applications that we are particularly interested in: driver's safety and patient's telemedicine (Lisetti et al. 2001). We believe this representation will prove useful because we can identify a finite set of possible emotional states, with their corresponding probabilities, in well-known contexts. The probability that a depressed patient goes from a state of depression to feeling ecstatic from one moment to the next is very low. On the other hand, we can expect a frustrated driver to become increasingly angry, if the stress of driving in New York City is added to some existing stress. Once enraged, his or her chances to move into a calm or contented state should also be expected to be very low. This is the intuition behind our model, which will then make it possible to predict and model humans or agents in a specific set of circumstances.

Internal Beliefs as Inputs

Finally, we want to point out that an individual's emotions can change in regard to an event, and these changes may be the result of their own efforts, not simply the result of an independent process directed by external events or social rules. Emotional changes indeed occur as a result of a number of processes, involving emotion dynamics rather than simply outside circumstances or the force of culture.

For example, one may experience guilt about being angry, or depression about feeling responsible. Emotional patterns can be transformed and changed as a result of external circumstances, or internal realizations, which in turn provoke further emotions. In our example, when and if one realizes that the experience of being angry was indeed justified, the feeling of being guilty about being angry vanishes. Similarly the feeling of being depressed about feeling responsible would vanish with the understanding of one's true lack of responsibility.

The first change would occur when one specific component of the anger emotion, namely *legitimacy*, is updated from its previous negative value to a positive one. The second change would occur when one updates the

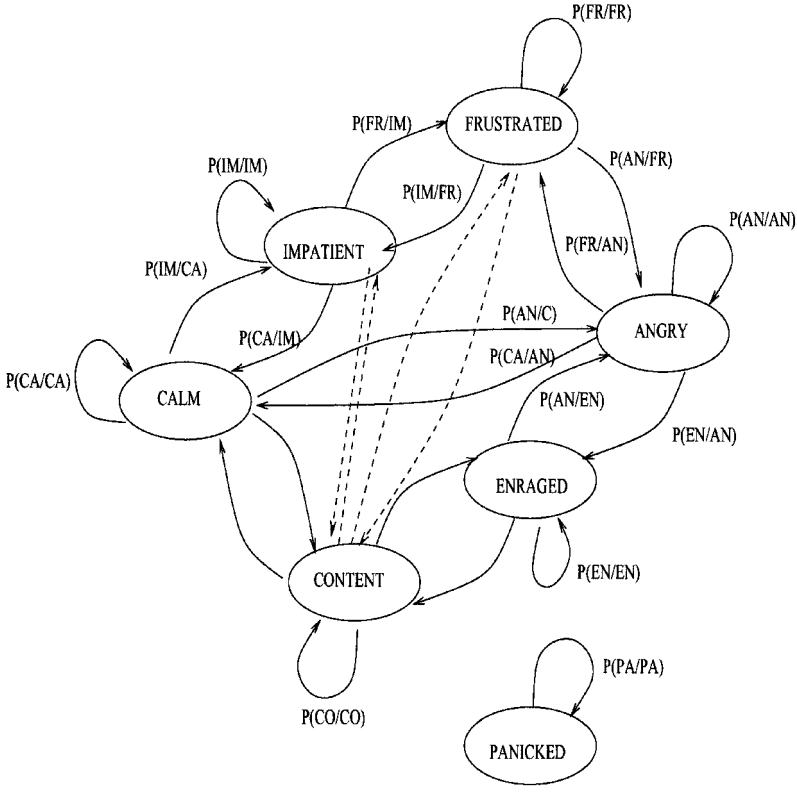

FIGURE 4. Driver's Markov Model.

agency attribute of the responsible feeling from self to other. Another example of such a displacement of current emotions is found when love turns to duty.

CONCLUSIONS

We have discussed directions in which *rational*$_1$ intelligent agents can be modeled so as to include the role of affect in their decision making. We have proposed a formalization of these roles in terms of emotional transformations. We have also created an Affective Knowledge Representation (AKR) scheme which should prove useful for future work in this area. Finally we have sketched a Markov Model approach to emotional state dynamics which, we hope, will be expanded upon.

Much work in the new area of Affective Computing is still needed as we slowly exit the Dark Ages of Cartesianism. We hope this article has accomplished its goal of starting a discussion as how to fully reincorporate

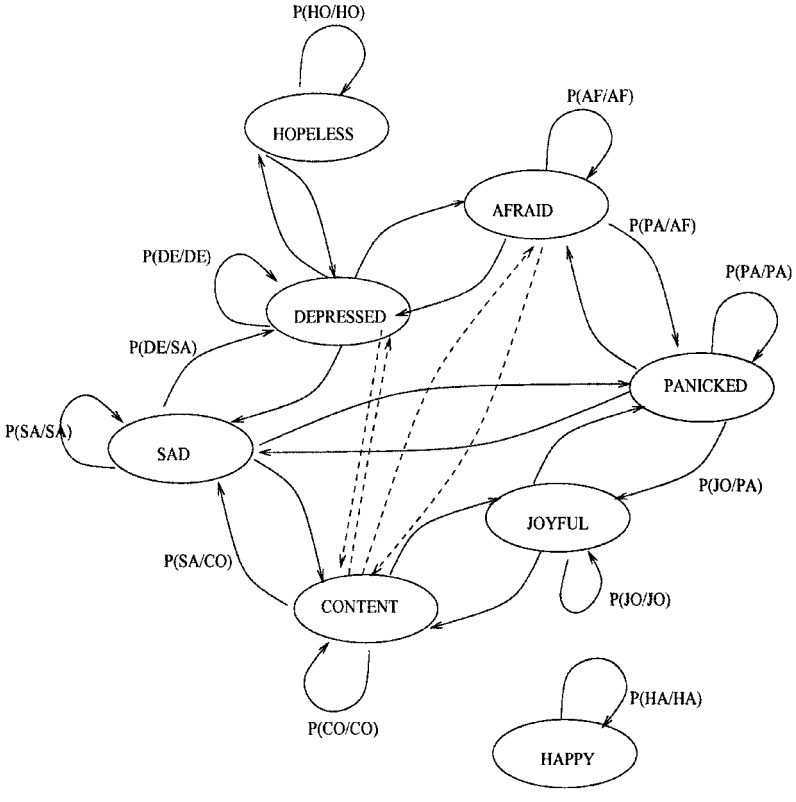

FIGURE 5. Depressed patient Markov Model.

the role of affect in *rationality*$_1$. We have talked about normative and descriptive models of decision making. We have said nothing about *prescriptive models* of decision making, which could potentially assist humans in making better decisions. We believe that prescriptive models will become possible to implement once we have arrived at more descriptive ones: for this, affect needs to be acknowledged.

NOTES

1. Many of the ideas exposed in this article originated from discussions which took place when Lisetti attended the Economics and Cognition Workshop 1996 at the Santa Fe Institute, New Mexico, USA; see (Lisetti 1996) and (Lisetti 1997b).
2. Applied to theories of the economy, these assumptions give rise to two different approaches: one based principally on *choice theory*, which deduces choice from their preferences, and the other is based on *revealed preference theory*, which from choice induces the preference which motivated the choice.
3. This article focuses principally on the role of emotions in decision making, but for a fuller list of other interactions of emotion with cognition (e.g., memory, communication, etc.), see the panel discussion in this issue.

4. Incidently, the influence of moods and emotions on perception is paralleled with similar findings about the influence of moods on memory: we recall an event better when we are in the same mood as when learning occured (Bower 1982).
5. It can be noted, however, that James had originally denied any content to emotion, describing them as purely physiological events (James 1884; 1894).
6. In natural organisms, valence and intensity are signaled by the activity of the autonomic nervous system, along the physiological dimension generated by the body proper, and do not necessarily involve the cognitive apparatus. They are available as early as the coarse level of affect (vs. emotion or mood). Hence, our model is consistent with Zajonc's theory on the primacy of affect (Zajonc 1980).

REFERENCES

Arnold, M. 1960. *Emotion and personality*. New York, NY: Columbia University Press.

Arthur, B. 1994. Inductive reasoning and bounded rationality. *Complexity in Economic Theory* 84(2): 406–411.

Arthur, B. 1995. Complexity in economic and financial markets. *Complexity* pages 20–25.

Barbalet, J. M. 1998. *Emotion, social theory, and social structure*. Cambridge, MA: Cambridge University Press.

Bohm, G., and H. Pfister. 1996. Instrumental or emotional evaluations: What determines preferences? *Acta Psychologica* 93:135–148.

Bower, G. 1982. Emotional influences in memory and thinking: Data and theory. In *Affect and cognition* eds., M. Clark and S. T. Fiske. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bower, G. H. 1981. Mood and memory. *American Psychologist* 36(2).

Breazeal, C. 1998. Infant-like social interactions between a robot and a human caretaker. *Adaptive Behavior*.

Brooks, R. 1987. Intelligence without representation. *MIT Artificial Intelligence Report*.

Burnett, M. S., and D. A. Lunsford. 1994. Conceptualizing guilt in the consumer decision-making process. *Journal of Consumer Marketing* 11(3).

Castelfranchi, C. 2000. Affective appraisal *vs.* cognitive evaluation in social emotions and interactions. In *Affect in interactions*, ed. A. Paiva. New York, NY: Springer Verlag.

Castelfranchi, C., F. de Rosis, and R. Falcone. 1997. Social attitudes and personalities in agents. In *Proceedings of the AAAI fall symposium series on socially intelligent agents*. Menlo Park, CA: AAAI Press.

Castelfranchi, C., and J.-P. Muller. 1993. From reaction to cognition. In Pages 3–9. *From reaction to cognition 5th European Workshop on Modelling an Agent in a Multi-Agent World* (*MAAMAW-93*) eds. C. Castelfranchi and J.-P. Muller, New York, NY: Springer Verlag.

Clancey, W. J. 1997. *Situated Cognition*. Cambridge, MA: Cambridge University Press.

Clark, A. 1996. *Being there: Putting brain, body and world together again*. Cambridge, MA: The MIT Press.

Collins, R. 1975. *Conflict sociology: Towards an explanatory science*. New York, NY: Academic Press.

Conati, C. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence* 16(7–8):555–575.

Connolly, T., L. D. Ordonez, and R. Coughlan. 1977. Regret and responsibility in the evaluation of decision outcomes. *Organizational behavior and human decision processes* 70(1):73–85.

Damasio, A. 1994. *Descartes' error*. New York, NY: Avon Books.

Daniels, K., and A. Guppy. 1997. Stressors, locus of control and social support as consequences of psychological well-being. *Journal of Occupational Health Psychology* 2:156–174

Darwin, C. 1872. *The expression of the emotions in man and animals*. London, England: Albermarle.

de Rosis, F. 2001. Preface to the special issue on affect in interaction. *User Modeling and User-Adapted Interaction* 11(4).

de Sousa, R. 1990. *The rationality of emotions*. Cambridge, MA: The MIT Press.

Derryberry, D., and D. Tucker. 1992. Neural mechanisms of emotion. *Journal of Consulting and Clinical Psychology* 60(3):329–337.

Descartes, R. 1637. *The philosophical works of Descartes (rendered into English by Haldane, E., and Ross, G.)* Cambridge, MA: Cambridge University Press.

Ekman, P. 1993. Facial expression and emotion. *American Psychologist* 48:384–392.
Ekman, P., R. Davidson, and W. Friesen. 1994. *The nature of emotion: Fundamental questions*. Oxford University Press.
Ekman, P., W. Friesen, and S. Tomkins. 1971. Facial affect scoring technique: A first validity study. *Semiotica* 3:37–58.
El-Nasr, M., T. Iorger, and J. Yen. 1999. Peteei: A pet with evolving emotional intelligence. In *Proceedings of the International Conference on Autonomous Agents*, pages 9–15.
Estrada, C., A. Isen, and M. Young. 1996. Positive affect improves creative problem-solving and in-fluences reported source of practice satisfaction in physicians. *Motivation and Emotion* 18: 285–299.
Evans, J., D. Over, and K. Manktelow. 1993. Reasoning, decision making and rationality. *Cognition*.
Frank, R. 1998. *Passions within reason: The strategic role of the emotions*. New York, NY: W. W. Norton.
Frijda, N., and J. Swagerman. 1987. Can computers feel? Theory and design of an emotional system. *Cognition and Emotion* 1(3):235–257.
Frijda, N. H. 1986. *The emotions*. New York: Cambridge University Press.
George, J. 1990. Personality, affect and behavior in groups. *Journal of Applied Psychology* 75:107–116.
Gmytrasiewicz, P., and E. Durfee. 2000. Rational coordination in multi-agent systems. In *Autonomous Agents and Multiagent Systems Journal*.
Gurtman, M., K. Martin, and N. Hintzman. 1990. Interpersonal reactions to displays of depression and anxiety. *Journal of Social and Clinical Psychology* 9:256–267.
Hatfield, E., J. Cacioppo, and R. Rapson. 1992. Primitive emotional contagion. In *Review of Personality and Social Psychology, Vol. 14: Emotion and Social Behavior*, ed. M. Clark. Newbury Park, CA: Sage.
Hayes-Roth, B., G. Ball, C. Lisetti, R. Picard, and A. Stern. 1998. Panel on affect and emotion in the user interface. In *Proceedings of the 1998 International Conference on Intelligent User Interfaces (IUI'98)*, pages 91–94. New York, NY: ACM Press.
Hirshleifer, J. 1987. On the emotions as guarantors of threats and promises. In *The latest on the best*, pages 307–326. Cambridge, MA: The MIT Press/Bradford.
Hudlicka, E., and M. McNeese. 2002. Assessment of user affect and belief states for interface adaptation: Application to air force pilot task. *User Modeling and User-Adapted Interaction* 12:1–47.
Hutchins, E. 1995. *Cognition in the *wild.* Cambridge, MA: The MIT Press.
Izard, C. E. 1977. *Human emotions*. New York, NY: Plenum Press.
James, W. 1884. What is an emotion? *Mind* 9:188–205.
James, W. 1894. The physical basis of emotion. *Psychological Review* 1:516–529.
James, W. 1897. *The will to believe and other essays (The sentiment of rationality)*. Dover Publications.
Johnson-Laird, P., and K. Oatley. 1989. The language of emotions: An analysis of a semantic field. *Cognition and Emotion* 3:81–123.
Johnson-Laird, P., and K. Oatley. 1992. Basic emotions, rationality, and folk theory. *Cognition and Emotion* 6(3/4):201–223.
Johnson-Laird, P., and E. Shafir. 1993. The interaction between reasoning and decision-making: An introduction. *Cognition* 49:1–9.
Koller, D., and A. Pfeffer. 1998. Probabilistic frame-based systems. In *Proceedings of the AAAI National Conference on AI*.
Kolm, S. 1986. The Buddhist theory of "no-self." In *The Multiple Self*, ed. J. Elster. Cambridge, England: Cambridge University Press.
Korzybski, A. 1933. *Science and sanity*. Lakeville: The International Aristotelian Library.
Lakoff, G. 1987. *Women, fire and dangerous things: What categories reveal About the mind*. Chicago, IL: University of Chicago Press.
Lane, D., and R. Maxfield. 1995. Foresight, complexity, and strategy. *Santa Fe Institute Studies in the Sciences of Complexity* 95-12-106.
Leventhal, H., and K. Scherer. 1987. The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion* 1(1):3–28.
Leventhal, H., and A. Tomarken. 1986. Emotion: Today's problems. *Annual Review of Psychology* 37: 565–610.
Levi-Strauss, C. 1978. *Myth and meaning*. Schoken Books.

Lisetti, C. L. 1996. Emotions-embodying-action. *Paper Presented at the Santa Fe Institute Economics and Cognition Workshop*.

Lisetti, C. 1997a. Motives for intelligent agents: Computational scripts for emotion concepts. In *Proceedings of the Sixth Scandinavian Conference on Artificial Intelligence* (*SCAI'97*), ed. G. Grahne, pages 59–70. Amsterdam, Holland: IOS press.

Lisetti, C. L. 1997b. Economics, cognition and the role of emotions in decision-making. *Stanford PDP Research Group Technical Report PDP-TR-002*.

Lisetti, C. 2002. Personality, affect and emotion taxonomy for socially intelligent agents. In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Symposium Conference* (*FLAIRS'*02). Menlo Park, CA: AAAI Press.

Lisetti, C. MOUE: A model of a user's emotions. *AI Communications*, in press.

Lisetti, C., and N. Bianchi-Berthouze. 2002. Modeling multimodel expression of user's affective subjective experience. *User Modeling and User-Adapted Interaction International Journal* 12(1):49–84.

Lisetti, C., M. Douglas, and C. LeRouge. 2001. Intelligent affective interfaces: A user-modeling approach for telemedicine. In *Proceedings of the First International Conference on Universal Access in Human-Computer Interaction* (*UAHCI'01*). Elsevier Science Publishers B. V.

Lisetti, C. and P. Gnytrasiewicz. 2000. Decisions, decisions... and the role of emotions in the process: A formal theory. In *Proceedings of the AAAI Fall Symposium Series on Socially Intelligent Agents* (*also AAAI Technical Report FS-00-04*). Menlo Park, CA: AAAI Press.

Lisetti, C. and D. Schiano. 2000. Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence, and cognitive science intersect. *Pragmatics and Cognition* 8(1):185–235.

Lisetti, C. L., D. E. Rumelhart, and M. Holler. 1998. An environment to acknowledge the interface between affect and cognition. In *Proceedings of the AAAI Spring Symposium Series on Intelligent Environment* (*also AAAI Technical Report SS-98-02*). Menlo Park, CA: AAAI Press.

Loewenstein, G. 1996. Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes* 65(3):272–292.

Lutz, C. 1985. Ethnopsychology compared to what? Explaining behavior and consciousness among the Ifaluk. In *Person, self, and experience: Exploring the pacific ethnopsychologies*, Berkeley, CA: eds. G. White and J. Kirkpatrick. University of California Press.

Mandler, G. 1975. *Mind and emotion*. New York, NY: Wiley.

Mellers, B. A., A. Schwartz, K. Ho, and I. Ritov. 1997. Decision affect theory: Emotional rections to the outcomes of risky options. *Psychological Science* 8(6):423–429.

Michaud, F., P. Prijanian, J. Audet, and D. Letourneau. 2000. Artificial emotions and social robotics. In *Proceedings of the International Symposium on Distributed Autonomous Robotic Systems*.

Minsky, M. 1975. A framework for representing knowledge. In *The Psychology of Computer Vision* eds. P. Winston and R. Brown. New York, NY: McGraw-Hill.

Moffat, D. 1997. Personality parameters and programs. In *Creating personalities for synthetic actors: Towards autonomous personality agents* (*LNAI 1195*), eds. R. Trappl and P. Petta, pages 120–165. Berlin Germany: Springer Verlag.

Munz, D. T. Huelsman, T. Konold, and J. McKinney. 1996. Are there methodological and substantive roles for affectivity in job diagnostic survey relationships? *Journal of Applied Psychology* 81:795–805.

Murphy, R., C. Lisetti, L. Irish, R. Tardif, and A. Gage. Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation* (in press).

Newell, A., and H. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Ortony, A., G. Clore, and A. Collins. 1988. *The cognitive structure of emotions* Cambridge, England: Cambridge University Press.

Ortony, A., and T. J. Turner. 1990. What's basic about basic emotions. *Psychological Review* 97:315–331, 1990.

Paiva, A. editor. 2000. *Affect in interactions*. Berlin, Germany: Springer Verlag.

Pfeifer, R. 1998. Artificial intelligence models of emotion. In *Cognitive perspectives on emotion and motivation*, eds. V. Hamilton, G. Bower, and N. Frijda, pages 287–320. Dordrecht, Netherlands: Kluwer Academic Publisher.

Pfister, H., and G. Bohm. 1992. The function of concrete emotions in rational decision making. *Acta Psychological* 80:199–211.

Picard, R., E. Vyzas, and J. Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(10):1175–1191, 2001.

Picard, R. W. 1997. *Affective computing*. Cambridge, MA:The MIT Press.

Plutchik, R. 1980. *Emotion theory, research and experience, volume 1: theories of emotion*. Academic press.

Rorty, R. 1979. *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press.

Roseman, A., A. Antoniou, and P. Jose. 1996. Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion* 10(3):241–277.

Rumelhart, D. E., and C. L. Lisetti. 1996. Emotions and consciousness: A connectionist approach. *Journal of Consciousness Studies*. (*Consciousness Research Abstracts: Toward a Science of Consciousness*).

Rumelhart, D. E., and J. L. McClelland. 1996. *Parallel distributed processing: Explorations in the microstructures of cognition, volume 1: foundations*. Cambridge, MA: The MIT Press/A Bradford Book.

Simon, H., 1967. Motivational and emotional controls of cognition. *Psychological Review* 1:29–39.

Sloman, A. 1990. Motives, mechanisms, and emotions. In *The Philosophy of Artificial Intelligence* ed. M. Boden. Oxford England: Oxford University Press.

Sloman, A., and M. Croucher. 1981. Why robots will have emotions. In *Proceedings of the Seventh IJCAI Vancouver, B.C.*, pages 197–202. San Mateo, CA: Morgan-Kaufmann.

Tomkins, S. S., and C. Izard. 1966. *Affect, cognition, and personality: Empirical studies*. London: Tavistock.

Tsiros, M. 1998. Effect of regret on post-choice valuation: The case of more than two alternatives. *Organizational Behavior and Human Decision Processes* 76(1):48–69.

Tversky, A. and D. Kahneman. 1986. Rational choice and the framing of decisions. *Journal of Business* 59(4):251–278.

Varela, F. 1989. *Autonomie et Connaissance*. Paris: Editions du Sevil.

Varela, E. Thompson, and E. Rosch. 1991. *The embodied mind: Cognitive science and human Experience*. Cambridge MA: The MIT Press.

Velasquez, J. 1999. From affect programs to higher cognitive emotions: An emotion-based control approach. In *Proceedings of the Emotion-Based Agent Architecture at the International Conference on Autonomous Agents*.

Wierzbicka, A. 1992. Defining emotion concepts. *Cognitive Science* 16:539–581.

Williams, J., F. Watts, C. McLeod, and A. Mathews. 1996. *Cognitive psychology and emotional disorders, 2nd Ed*. Chichester England: Wiley.

Zajonc, R. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151–175.

Zajonc, R. 1984. On the primacy of affect. *American Psychologist* 39:117–124.

Zeelenberg, M., W. Can Dijk, J. van Der Pligt, A. Manstead, P. Van Empelen, and D. Reinderman. 1998. Emotional reactions to the outcomes of decisions: The role of counterfactual thought in the experience of regret and disappointment. *Organizational Behavior and Human Decision Processes* 75(2):117–141.