Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Cognitive Systems Research 5 (2004) 269–306

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A domain-independent framework for modeling emotion

Action editor: Paul Thagard

Jonathan Gratch [a,*], Stacy Marsella [b]

[a] *Institute for Creative Technologies, University of Southern California, 13274 Fiji Way, Marina del Rey, CA 90292, USA*
[b] *Information Sciences Institute, University of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292, USA*

Received 12 November 2003; accepted 13 February 2004
Available online 13 April 2004

Abstract

In this article, we show how psychological theories of emotion shed light on the interaction between emotion and cognition, and thus can inform the design of human-like autonomous agents that must convey these core aspects of human behavior. We lay out a general computational framework of appraisal and coping as a central organizing principle for such systems. We then discuss a detailed domain-independent model based on this framework, illustrating how it has been applied to the problem of generating behavior for a significant social training application. The model is useful not only for deriving emotional state, but also for informing a number of the behaviors that must be modeled by virtual humans such as facial expressions, dialogue management, planning, reacting, and social understanding. Thus, the work is of potential interest to models of strategic decision-making, action selection, facial animation, and social intelligence.

> The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions. Marvin Minsky (Minsky, 1986, p. 163)

> In every art form it is the emotional content that makes the difference between mere technical skill and true art. The Illusion of Life: Disney Animation (Thomas & Johnston, 1995, p. 473)

> You don't get emotions by manipulating 0s and 1s. John Searle (Searle, 2002)

1. Introduction

Emotions play a powerful, central role in our lives. They impact our beliefs, inform our decision-making and in large measure guide how we adapt our behavior to the world around us. While most apparent in moments of great stress, emotions sway even the mundane decisions we face in everyday life (Clore & Gasper, 2000; Damasio, 1994). Emotions also infuse our social relationships. Our interactions with each other are a source of many of our emotions and we have developed both a range of behaviors that communicate emotional information as well as an ability to recognize the emotional arousal in others. By

[*] Corresponding author.
E-mail addresses: gratch@ict.usc.edu (J. Gratch), marsella@isi.edu (S. Marsella).
URLs: www.ict.usc.edu/~gratch, www.isi.edu/~marsella.

1389-0417/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2004.02.002

virtue of their central role and wide influence, emotion arguably provides the means to coordinate the diverse mental and physical components required to respond to the world in a coherent fashion (Cosmides & Tooby, 2000).

The goal of our research is to create a general computational model of the mechanisms underlying human emotion that accounts for this range of phenomena. Although such a model can ideally inform our understanding of human behavior, we see the development of computational models of emotion as a core research focus for artificial intelligence that will facilitate advances in the large array of computational systems that model, interpret or influence human behavior. On the one hand, modeling applications must account for how people behave when experiencing intense emotion including disaster preparedness (e.g., when modeling how crowds react in a disaster (Silverman, 2002)), training (e.g., when modeling how military units respond in a battle (Gratch & Marsella, 2003)), and even macro-economic models (e.g., when modeling the economic impact of traumatic events such as 9/11 or the SARS epidemic). On the other hand, applications presume the ability to correctly interpret the beliefs, motives and intentions underlying human behavior (such as tutoring systems, dialogue systems, mixed-initiative planning systems, or systems that learn from observation) and could benefit from a model of how emotion motivates action, distorts perception and inference, and communicates information about mental state. Finally, emotions play a powerful role in social influence, a better understanding of which would benefit applications that attempt to shape human behavior, such as psychotherapy applications (Marsella, Johnson, & LaBore, 2000, 2003; Rothbaum et al., 1999), tutoring systems (Lester, Stone, & Stelling, 1999; Ryokai, Vaucelle, & Cassell, 2003; Shaw, Johnson, & Ganeshan, 1999), or marketing applications (André, Rist, Mulken, & Klesen, 2000; Cassell, Bickmore, Campbell, Vilhjálmsson, & Yan, 2000). A general computational account of emotion could benefit this wide range of applications, but additionally, by unifying them within a common conceptually framework, it would facilitate the juxtaposition of findings from across these disparate applications, ultimately improving our understanding of human emotion and the mechanisms which underlie it.

A secondary, more speculative motivation for building general models of emotion is that they may give insight into building models of intelligent behavior *in general*. Several authors have argued that emotional influences that seem irrational on the surface have important social and cognitive functions that would be required by any intelligent system (Damasio, 1994; Minsky, 1986; Oatley & Johnson-Laird, 1987; Simon, 1967; Sloman & Croucher, 1981; Lisetti & Gmytrasiewicz, 2002). For example, social emotions such as anger and guilt may reflect a mechanism that improves group utility by minimizing social conflicts, and thereby explains peoples "irrational" choices in social games such as prison's dilemma (Frank, 1988). Similarly, "delusional" coping strategies such as wishful thinking may reflect a rational mechanism that is more accurately accounting for certain social costs (Mele, 2001).

In this article, we layout our current progress towards a unified model that can simulate human emotional responses but also inform the debate on the general adaptive value of emotional reasoning. We show how certain psychological theories shed light on the processes underlying emotion and its influence on cognition, and thus can serve as a basis of a computational model. Our goal is to model the range of human emotions, as well as their dynamics: the depression after a relationship that breaks up that turns into anger at the former partner, fear that transforms into anger at the causes of that fear. Although modeling such complex emotions and emotional dynamics may seem implausible at first, significant advances in emotion psychology, beginning with work of Arnold (1960), shed considerable light on the design of emotional virtual humans. This work characterizes emotion as the result of underlying mechanisms including *appraisal*, that evaluates an organism's circumstances, and *coping*, that guides the response to this assessment (Frijda, 1987; Lazarus, 1991; Ortony, Clore, & Collins, 1988; Peacock & Wong, 1990; Scherer, 1984; Scherer, Schorr, & Johnstone, 2001). These *appraisal theories* argue that appraisal and coping not only underlie emotional behavior, but play an essential role in informing cognition, often in ways not considered by traditional models of intelligence.

In particular, we lay out a general computational framework of appraisal and coping as core reasoning components for human-like autonomous agents. In this, we extend prior computational models of appraisal by incorporating a general process model of the influences between cognition and appraisal, and provide what we believe to be the first computational account of coping. We then discuss a detailed implementation based on this framework. By recasting appraisal theory in terms of the algorithms and data-structures that underlie many autonomous agent systems, we argue that this approach provides an important control construct for linking various agent components in a tighter and more coherent fashion. We illustrate this point by showing how this viewpoint facilitates the integration of such disparate reasoning modules as perception, planning, and dialogue management into a coherent appraisal of the agent's relationship to its environment. Beyond illustrating this integration, a model of appraisal is useful in of itself, not only for deriving emotional state, but also for informing a number of the behaviors that must be modeled by human-like agents such as facial expressions, dialogue management, planning, reacting, and social understanding. These points are realized in an implemented system, which has been applied to a significant real world problem.

Section 2 summarizes appraisal theory, lays out the requirements a model must satisfy, and sketches the general outlines of a computational approach. Section 3 discusses related work. Section 4 describes EMA, a computational model of appraisal and coping consistent with this reinterpretation. Section 5 describes an application of EMA to the problem of modeling human emotional behavior in a virtual reality training system. Section 6 discusses the implications of our model for general models of intelligent behavior and summarizes the outstanding issues and limitations of our current approach.

2. Theoretical frame work

A challenge in working towards a general computational model of emotion is the considerable controversy as to what form such a theory should take. There is no universally accepted definition of emotion, nor is there general consensus on the range of phenomena that constitute the domain of emotion research. Some theories argue for a set of distinct emotions with neurological correlates and well circumscribed effects (Ekman, 1972; LeDoux, 1996), whereas others argue that emotions are epiphenomenal, simply reflecting the interaction of underlying processes. Some theories argue that emotions arise from physiological processes in the body that subsequently impact cognition (e.g., James–Lange theory), whereas other argue that the causality is reversed (Lazarus, 1991), or a combination of the two (Damasio, 1994). The distinction between emotion and other related constructs such as feeling, mood or personality is similarly murky. For example, some theories distinguish these constructs simply by their behavioral time course: emotions are behavioral dispositions that persist for seconds or minutes, moods are states that have similar effects but over a longer time course of hours or days, while personality traits reflect relatively stable behavioral tendencies.

We approach the problem of modeling emotion from a symbolic artificial intelligence perspective that shapes our interpretation of these competing theories. Rather than solely seeking to provide definitions for certain emotional states, we are interested in modeling the mechanisms that underlie emotional behavior and its influence on mental processing. Our focus is also on "broad agents" that integrate a number of symbolic reasoning processes such as planning, acting, natural language communication and user modeling into a single system (Bates, Loyall, & Reilly, 1991) and the role emotion processing may play in this integration. In particular, we have addressed the problem of building *virtual humans*, software entities that look and act like people, but live in simulated graphical environments and can freely interact with humans immersed in the environment (Gratch et al., 2002).

Our symbolic focus is a natural fit for appraisal theories that emphasize the tight relationship between emotion and symbolic reasoning, though it deemphasizes the bodily sources and consequences of emotions argued by many theorists (Zajonc, 1980). As a consequence, our current model is best suited to researchers interested in more symbolic systems and deliberative reasoning. It is less suited to researchers interested in lower-level processes such as perception and non-deliberate reasoning (e.g., reactions), though we discuss in Section 6.7 how these views might be reconciled.

Our focus is also on process rather than surface behavior. We desire a model that explains, for example how emotion might arise from an agent's reasoning processes and subsequently impact decision-making. From this perspective, the specific definition of emotional terms such as "joy" or "fear" are less important than the processes that underlie them. In fact, in this article we will largely avoid specific emotion terms and instead focus on specific mechanisms such as appraisal and coping. As a consequence, our model is best suited to researchers interested in the relationship of emotion to cognitive processes, as well as the adaptive function that these processes may provide. Appraisal theory most directly addresses the issues raised from this perspective, and thus serves as the theoretical basis for our model.

Appraisal theory serves as the conceptual basis for our work, but this psychological theory is insufficiently precise to serve as a specification of a computational model. For this, we recast the theory in terms of artificial intelligence methods and representations. This section lays out this basic theory and then considers what implications it has for agent design and what artificial intelligence techniques best handle the constraints that this theory imposes. Smith and Lazarus (1990) cognitive–motivational–emotive system, illustrated in Fig. 1, is representative of contemporary appraisal theories.[1] Emotion is conceptualized as a two-stage control system. Appraisal characterizes the relationship between a person and their physical and social environment, referred to as the *person–environment relationship* and coping recruits resources to repair or maintain this relationship. Behavior arises from a close coupling of cognition, emotion and coping responses: cognitive processes serve to build up an individual's interpretation of how external events relate to their goals and desires (the person–environment relationship); Appraisal characterizes this interpretation in terms of a number of abstract features that are useful for guiding behavior; Coping draws on these characterizations to alter the person–environment relationship, by motivating actions that change the environment (problem-focused coping), or by motivating changes to the interpretation of this relationship (emotion-focused coping).

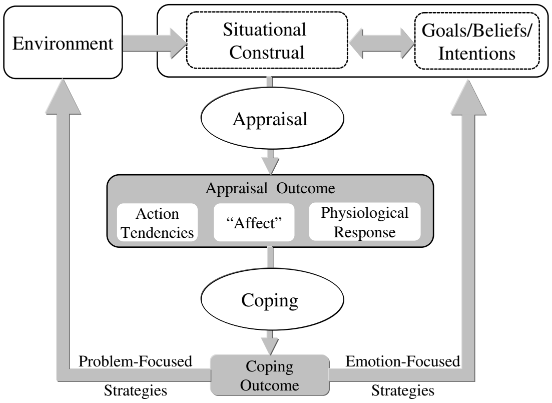

Fig. 1. The cognitive–motivational–emotive system. Adapted from Smith and Lazarus (1990).

In developing a model of emotion for virtual humans, one must decide at what level to model behavior. Systems that mimic human behavior have been traditionally divided into cognitive models that mimic underlying mental processes versus techniques that attempt to replicate surface behavior, independent of the accuracy of the underlying processes (much of the work on virtual humans has actually focused on the even more modest goal of simply producing "believable" behavior). The problem of modeling realistic emotional and non-verbal behavior, however, has

[1] In this article we present cognitive appraisal theory as a single, unified theory, and though theorists largely agree in the abstract, there are important differences between individual models and their terminology. For example, Roseman argues for a much tighter linkage between appraisal and coping, referring to coping strategies as "emotivational goals" that are associated with specific emotions (Roseman, Wiest, & Swartz, 1994), whereas other theories view coping as outside the scope of appraisal theory. See (Ellsworth & Scherer, 2003) for an excellent review of contemporary appraisal theory.

blurred this traditional distinction. For example, most people accept as obvious that facial expressions, gestures and choice of action reflect the dynamics of underlying mental processes and actors and animators exploit this assumption to great effect. We adopt this assumption and, more strongly, argue that, to convey realism in an *interactive* setting, the external manifestation of agent behavior must be linked to the agent's internal processing. This, by necessity, constrains these internal processes towards greater realism. In realizing these constraints, our research methodology relies on psychological theories to inform behavioral requirements and suggest high-level process models, but uses traditional artificial intelligence techniques to make these theories concrete. We feel this approach is essential for building working systems but may also be of some interest to psychologists. For example, even if many artificial intelligence algorithms are psychologically implausible as process models, they can be viewed as initial approximations and could help concretize psychological process theories. However, from the perspective of producing realistic behavior, the plausibility of these techniques depends only on their impact on external behavior.

2.1. The cognitive–motivational–emotive system

Smith and Lazarus' theory organizes behavior around two basic processes, appraisal (which characterizes the person's relationship with their environment), and coping (which suggests strategies for altering or maintaining this relationship). Cognition informs both of these processes. It informs appraisal by constructing mental representations of how events relate to internal dispositions such as goals. It informs coping by suggesting and exploring strategies for altering or maintaining the person–environment relationship. The role of cognition in these theories has led some to criticize them as excessively deliberate or "cold" models of emotion that may be appropriate for reasoning about emotion but don't actually inform external behavior. However, appraisal should not be construed as a deliberate process in itself, but rather a reflexive assessment of the current mental state, which may or may not have been elaborated by deliberation (Lazarus, 2001, pp. 178–180). Thus, appraisal does not require sophisticated reasoning, but is able to exploit the output of whatever reasoning has been performed, sophisticated or not.[2]

2.1.1. Appraisal and appraisal variables

Appraisal theories posit that events do not have significance in of themselves, but only by virtue of their interpretation in the context of an individual's beliefs, desires, intentions and abilities. For example, the outcome of the latest presidential election might inspire joy, anger or indifference, depending on how the candidate's policies are believed to impact one's goals. Appraisal theories argue for the central role of *appraisal variables* (or sometimes called appraisal components or appraisal dimensions) in characterizing this interpretation. These are essentially criteria along which the significance of events can be judged. Table 1 summarizes key variables identified by a number of appraisal theories. As a computational construct, appraisal variables can provide considerable traction towards developing domain-independent computational models of emotion, motivation, and behavior because they focus the messy details of cognitive processes into a tractable number of domain-independent mediating concepts that can subsequently inform behavior.

Though originally developed to explain emotion, appraisal variables appear related to a wide range of psychological concepts, and are correlated with differences in decision-making, action selection, coping, personality and culture (Costa, Somerfield, & McCrae, 1996; Penley & Tomaka, 2002; Scherer et al., 2001) and a general mechanism for deriving such variables could aid in modeling multiple aspects of human behavior. For example, appraisal variables appear to mediate an organism's response to stimuli: rather than associating responses directly with perceptual features, as in reactive planning systems (Agre & Chapman, 1987), responses seem organized around the organisms appraised interpretation of events. For

[2] In fact, appraisals in humans may arise from multiple processes, some rapid and perceptual based, e.g., see (Smith & Kirby, 2000).

Table 1
Appraisal variables

Relevance		Does the event require attention or adaptive reaction
Desirability		Does the event facilitate or thwart what the person wants
Causal attribution	Agency	What causal agent was responsible for an event
	Blame and credit	Does the causal agent deserve blame or credit
Likelihood		How likely was the event; how likely is an outcome
Unexpectedness		Was the event predicted from past knowledge
Urgency		Will delaying a response make matters worse
Ego involvement		To what extent does the event impact a person's sense of self (social esteem, moral values, cherished beliefs, etc.)
Coping potential	Controllability	The extent to which an event can be influenced
	Changeability	The extent to which an event will change of its own accord
	Power	The power of a particular agent to directly or indirectly control an event
	Adaptability	Can the person live with the consequences of the event

example, people's strategic choice of problem-focused vs. emotion-focused responses seems influenced by whether subjects appraise a situation as threatening or challenging (Peacock & Wong, 1990) and (Frijda, 1987) has demonstrated associations between reactive behaviors (flight vs. fight) and configurations of appraisal variables. In terms of lower-level expressive behaviors, studies have indicated that appraisals lead to specific response patterns in facial expression (Scherer, 1984; Smith & Scott, 1997), verbal expression (Banse & Scherer, 1996) and physiological response (Kirby & Smith, 1996). Several recent theories of personality have also suggested that appraisal variables could serve as the basis of models of personality. These theories point to correlations between a person's traits and biases in the way events are appraised. For example, extraversion has been positively correlated with perceived accountability and control (Penley & Tomaka, 2002). Thus, a general computational model of appraisal is of potential interest to models of strategic decision-making, action selection, facial animation, and personality.

Beyond modeling the significance of events to one's self, appraisal variables also seems to play an important role in mediating social relationships. People readily appraise how events impact other individuals and use these appraisals to guide social actions. For example, if a person believes his actions have harmed another, he may be moved to redress his wrong, even in the absence of any evidence that the other actually shares this interpretation. Such "anticipatory guilt" seems to play a key role in enforcing social norms like fairness (Frank, 1988). People seem to use social appraisals to inform the interpretation of other individual's behavior. For example, an ambiguous response to an event may be interpreted differently depending on what appraisals we attribute to the responder. Here, we argue for using the same computational model to assess an agent's own appraised relationship to the environment as well as the imagined relationship between other agents and their environment. By assessing the imagined beliefs and preferences of other entities and appraising events from their perspective, the model allows an agent to use these appraisals to influence social actions and interpretations. Thus, a model of appraisal is also of potential use to researchers interested in models of social-intelligence.

2.1.2. Coping

Coping determines how one responds to the appraised significance of events. These events may be in the past, in the present or in the future. Thus people need to cope with current events as well as cope with what they believe will happen in the future or their memories of past events. Evidence suggest that people are motivated to respond to events differently depending on how they are appraised (Peacock & Wong, 1990). These studies suggest, for example, that events appraised as undesirable but controllable motivate people to develop and execute plans to reverse these circumstances. On the other hand, events appraised as uncontrollable lead people towards escapism or resignation. Computational approaches that model this motivational function have largely focused on the former, using emotion or appraisal to guide external action, however psychological theories characterize coping more broadly. In addition to acting on the environment, which has been termed *problem-focused coping*, people employ inner-directed strategies for dealing with strong emotions, termed *emotion-focused coping* (Lazarus, 1991). Emotion-focused coping works by altering one's interpretation of circumstances, for example, by discounting a potential threat or abandoning a cherished goal. In addition to organizing coping strategies into these two broad categories (sometimes researchers add *suppression* as a third separate category), coping researchers have distinguished numerous techniques people use to cope. Table 2 illustrates the variety of distinct ways people may cope with their circumstances, adapted from (Carver, Scheier, & Weintraub, 1989).

Coping relies on appraisal to identify significant features of the person–environment relationship and to assess the potential to maintain or overturn these features (coping potential). Based on this assessment, coping selects amongst competing strategies to alter this relationship. For example, if one feels guilty about causing a traffic accident, one may be motivated to redress the wrong (problem-focused coping) or alternatively, shift-blame to the other driver (emotion-focused coping). Coping relies on a range of cognitive process to realize these strategies. So, whereas coping may form the intention to redress the wrong, cognition must still devise a particular plan of attack. The ultimate effect of these strategies is a change in the person's interpretation of their relationship with the environment, which can lead to new appraisals. Thus, coping, cognition and appraisal are tightly coupled, interacting and unfolding over time (Lazarus, 1991; Scherer, 1984): an agent may interpret a situation as threatening (appraisal), which in turn motivate the shifting of blame (coping), which leads to anger (re-appraisal).

Note the distinction between problem-directed and emotion-directed is not crisp. In particular, much of what counts as problem-focused coping in the psychological literature has an inner-directed aspect. For example, one might form an intention to achieve a desired state – and feel better as a consequence – without ever acting on the intention. Thus, by performing cognitive acts like planning, one can improve ones interpretation of circumstances without actually changing the physical environment. One of our goals is to use computational models as a means to move towards more precise ways of distinguishing coping strategies.

2.2. A computational perspective

A central tenet in appraisal theories is that appraisal and coping center around a person's *interpretation* of their relationship with the

Table 2
Some common coping strategies

Problem-focused coping	Active coping: taking active steps to try to remove or circumvent the stressor Planning: thinking about how to cope. Coming up w/ action strategies Seeking social support for instrumental reasons: seeking advice, assistance, or information
Emotion-focused coping	Suppression of competing activities: put other projects aside or let them slide. Restraint coping: waiting till the appropriate opportunity. Holding back Seeking social support for emotional reasons: getting moral support, sympathy, or understanding. Positive reinterpretation and growth: look for silver lining; try to grow as a person as a result. Acceptance: accept stressor as real. Learn to live with it Turning to religion: pray, put trust in god (assume God has a plan) Focus on and vent: can be function to accommodate loss and move forward Denial: denying the reality of event Behavioral disengagement: Admit I cannot deal. Reduce effort Mental disengagement: Use other activities to take mind off problem: daydreaming, sleeping Alcohol/drug disengagement

environment. This interpretation is constructed by cognitive processes, summarized by appraisal variables and altered or reinforced by coping responses. In recasting appraisal theory in computational terms, we must consider what types of representations and reasoning techniques would support such a process. Any model that claims to support the full range of appraisal variables and coping strategies listed in the preceding tables must minimally satisfy the following requirements:

- To capture the constructive and interpretative nature of these processes and to model their dynamic unfolding over time through the tightly coupled interaction of cognition, appraisal and coping, the model must explicitly represent intermediate knowledge states that may be appraised and augmented by further inference.
- To reason about relevance and desirability, the model must represent preferences over outcomes.
- To make causal attributions, the model must represent some notion of causality and agency (e.g., one might blame a Serbian nationalist, through a series of intervening events, for causing the First World War). Blame may be assigned to past or future events, so the model must represent past as well as future causal relations.
- To reason about likelihood, unexpectedness and changeability, the model must represent causal factors influencing events, future possible outcomes and interactions between possible outcomes (e.g., does the plan to achieve goal A interfere with the plan to achieve goal B?).
- To reason about the urgency, the model must represent temporal constraints, event duration, and, potentially, partial goal achievement as a function of time.
- To reason about controllability, the model must represent the extent to which events can be controlled (e.g., how robust is my plan?).
- To reason about social power, the model must have some representation of coercive relationships between agents such as representing different agent's sphere of influences or organizational hierarchies.
- To reason about adaptability and to support emotion focused coping strategies, the model must be open to subjective reinterpretation (e.g., represent subjective rather than "true" beliefs).
- To reason about ego-involvement, the model must support some notion of how "central" a desire is to the agent's self-concept.

What this list suggests is that an intelligent system that hopes to mimic the generality of human emotional capabilities will have to integrate, as a minimum, these functions into a single architecture.[3] With the possible exception of ego-involvement, most of these requirements have been addressed in one form or another by computational systems. For example, preferences over outcomes typically represented through the use of goals and/or utility functions. Causal attributions typically involve reasoning about the beliefs and intents of other agents (did he intend to harm me or did he foresee the consequences of his actions), typically handled by logics of intention and belief. Likelihood or future representations are typically represented through plans, Bayesian networks or Markov chains. Reasoning about urgency typically involves some form of temporal reasoning and may involve temporal logics. Deriving an agent's sense of how much a situation will change or can be controlled involves reasoning about causality, typically handled by planning techniques, causal models (Pearl, 2002) or envisionments (de Kleer & Brown, 1982). Representational schemes that support subjective or adaptable beliefs about the world include the use of subjective probabilities (Russell & Wefald, 1989) or higher-order logics (Reiter, 1987).

To satisfy these requirements, we have found it most natural to build on plan-based causal representations, augmenting them with decision-theoretic planning techniques (Blythe, 1999; Boutilier, Dean, & Hanks, 1999) and with methods that explicitly model commitments to beliefs and intentions (Bratman, 1990; Grosz & Kraus, 1996; Pollack, 1990). Although neither of these techniques was developed to model realistic human

[3] Even this is too small for a general model as it reflects the cognitive bias of cognitive appraisal theories and deemphasizes the importance of lower-level perceptual and reactive processes. Of course, depending on how a specific application is crafted, some of these capabilities can be 'engineered away'.

behavior (indeed, there is considerable evidence that people violate the rules of classical decision-theory), taken together, they provide a first approximation of the type of reasoning that underlies appraisal and coping. They also greatly facilitate the design of working agent systems. Each technique satisfies a portion of the requirements, but only in combination are they sufficient to support a full model of appraisal theory. Plan representations provide a concise representation of the causal relationship between events and states, key for assessing the relevance of events to an agent's goals and for assessing causal attributions. Plan representations also lie at the heart of a number of autonomous agent reasoning techniques (e.g., planning, explanation, natural language processing). Decision-theoretic planning models take the additional step of combining plan reasoning with reasoning about uncertainty and the desirability of different outcomes, satisfying the requirements of reasoning about likelihood and desirability. They do not, however, model the notion of commitment to a belief or intention, key for forming attributions of blame or credit (which involve reasoning if the causal agent intended or foresaw the consequences of their actions) and for assessing the significance of events to other agents (as when I believe an outcome is good but I believe that you believe it to be bad). We draw on models of beliefs and intentions to provide this distinction (Bratman, 1990; Grosz & Kraus, 1996; Pollack, 1990), though these approaches by themselves fail to make a number of distinctions important for modeling appraisal variables (e.g., likelihood, desirability, unexpectedness, controllability, urgency, future expectancy). For example, they don't represent variable preferences over outcomes, alternative competing courses of action, temporal constraints, imperfect information or probabilistic effects. By combining these approaches, we have arrived at a set of representational distinctions and reasoning techniques that span the bulk of the requirements underlying appraisal and coping.

We use the term *causal interpretation* to refer to a particular instantiation of this amalgam of plans, beliefs, desires, intentions, probabilities and utilities mentioned above. This terminology emphasizes the importance of causal reasoning as well as the interpretative (subjective) character of the appraisal and coping processes. From an AI perspective, the causal interpretation refers to a current 'mental state' (Bratman, 1990; Pollack, 1990) and serves to inform and constrain subsequent inference. In the terminology of Smith and Lazarus, the causal interpretation is as a declarative representation of the current construal of the person–environment relationship.

Finally, appraisal is said to operate over "events" that impact the person–environment relationship and we must concretize the relationship between the causal interpretation, events, appraisal variables and emotions. Appraisal theories are short on details and somewhat contradictory on this point, on the one hand arguing that appraisal is some overall assessment of the person–environment relationship (which imposes a global perspective over a person's past, present and future circumstances), and on the other hand arguing that appraisal is triggered by specific physical events. This is further complicated by the interpretative nature of appraisal. Is an event something that is "out there" in the environment? Is it the interpretation of some external phenomena? Is it purely mental? See (Shaver, 1985) and (Lazarus, 1991) for a discussion. Following our representational commitments, we have found it most natural to model an agent's overall assessment as an aggregation of how the agent appraises individual events. We then define an event to be any physical action represented in the causal interpretation that is believed to facilitates or inhibits some (past, present or future) state with non-zero utility for the agent (or for other entities that the agent is aware of).[4] These various terms will be made

[4] This definition can be generalized. For example, it downplays the importance of predictability. An event might be interesting simply because it violates one's expectations, and thus serves as a learning opportunity. Though our model detects unexpected events, it does not attempt to attribute this to a defect in its knowledge. Thus, it cannot currently motivate this kind of learning. A second insufficiency is it precludes appraising mental or communicative acts directly. For example, one could appraise the implications of a communicative act but not the act itself (i.e., one couldn't blame the messenger). Similarly, one couldn't rue a bad decision.

concrete when we detail the representation of the causal interpretation in the following section.

To summarize, in our conceptualization of appraisal theory, the agent's causal interpretation is equated with the output and intermediate results of those reasoning algorithms that relate the agent to its physical and social environment. At any point in time, this configuration of beliefs, desires, plans, and intentions represents the agent's current view of the agent-environment relationship, an interpretation that may subsequently change with further observation or inference. We treat appraisal as mapping from domain-independent features of causal interpretation to individual appraisal variables. For example, an effect that overturns a desired goal would lead to distress fear. Multiple appraisals are aggregated into an overall emotional state that influences behavior. Coping directs control signals to auxiliary reasoning modules (i.e., planning, action selection, belief updates, etc.) to overturn or maintain features of the causal interpretation that lead to individual appraisals. For example, coping may resign the agent to the threat by abandoning the desire. Fig. 2 illustrates a reinterpretation of Smith and Lazarus' cognitive–motivational–emotive system consistent with this view. The causal interpretation could be viewed as a representation of working memory (for those familiar with psychological theories) or as a blackboard (for those familiar with blackboard architectures).

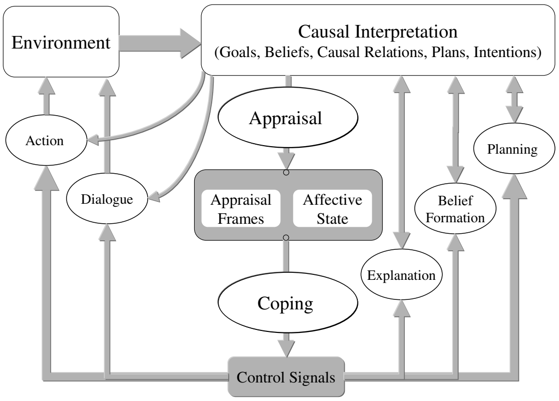

Fig. 2. Our computational instantiation of the cognitive–motivational–emotive system.

3. Related work

In laying the groundwork for a general computational accounting of appraisal and coping, our work relates to a number of past approaches. Although we are perhaps the first to provide an integrated computational account of coping, computer science researchers have made steady progress over the years in appraisal-based approaches to emotional expression and action selection. Beyond a model of coping, our work contributes primarily to the problem of developing general and domain-independent algorithms to support appraisal, and by extending the range of appraisal variables amenable to a computational treatment. Prior computational appraisal models have focused almost exclusively on the work of Ortony et al. (1988), whose work has not similarly dominated psychological thinking in recent years, so another contribution of our current work is to broaden the range of discourse on appraisal theories within the computational community.

Early computational models of appraisal focused on the mapping between appraisal variables and behavior and largely ignored how these variables might be derived, focusing on domain-specific schemes to derive their value variables. For example, Elliott (1992) Affective Reasoner, a computational realization of the theory of Ortony et al. (1988), required a number of domain specific rules to appraise events. A typical rule would be that a goal at a football match is desirable if the agent favors the team that scored. While useful as a first approximation, the reliance on such specific rules doesn't give much insight on how to make such inferences in general. Nor do they provide any insight on how to integrate appraisals with coping responses. The Affective Reasoner, following Ortony et al.'s, carves up appraisal variables differently than Table 1, in particular emphasizing the role of *standards* in attributing blame or credit. We return to this issue in Section 6.1. In terms of our model, a key contribution of Elliot's model is the use of explicit appraisal frames that characterize events in terms of specific appraisal variables, and the notion of appraising the same event from multiple perspectives (from the agent's own perspective, and the supposed perspective of other agents). We build on both of these notions.

More recent approaches have moved toward more abstract reasoning frameworks, largely building on traditional artificial intelligence techniques. For example, Neil Reilly (1996) EM algorithm associates desirability with probability of goal attainment in a reactive planning framework, although this system relied on domain-specific rules to derive the likelihood of threats or goal attainment. El Nasr, Yen, and Ioerger (2000) uses Markov-decision processes (MDP) to provide a very general framework for characterizing the desirability of actions and events. An advantage of this method is that it can represent indirect consequences of actions by examining their impact on future reward (as encoded in the MDP), but it retains the key limitations of such models: they can only represent a relatively small number of state transitions and assume fixed goals. MDPs are also ill suited to model the collaborative planning and conversations about plans that most virtual human applications have addressed. The closest computational approach to what we propose here is the Will architecture (Moffat & Frijda, 1995) that ties appraisal variables to an explicit model of plans (which capture the causal relationships between actions and effects), although they, also, did not address the issue of blame/credit attributions, or how coping might alter this interpretation. We build on these prior general models, extending them to provide a better characterization of causality and the subjective nature of appraisal that facilitates coping.

Besides appraisal-driven models of emotion, a number of researchers have explored communication-driven approaches. In these models, an emotional expression is chosen based on some desired impact it will have on the user. For example, Poggi and Pelachaud (2000) use facial expressions to convey the performative of a speech act, showing "potential anger" to communicate that the agent will be angry if a request is not fulfilled. Tutoring applications usually also follow a communication-driven approach, intentionally expressing certain emotions with the goal to motivate the student. An example includes the Cosmo system in which the selection and sequencing of emotive behaviors is driven by the agent's pedagogical goals (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000). For instance, a congratulatory act triggers a motivational goal to express admiration that is conveyed with applause. A disadvantage of communication-driven approaches is that, as emotional behaviors are not tied to a consistent appraised state, they can swing widely or seem insincere. Appraisal-based and communication-based approaches are complementary. Appraisal methods could improve the perceived coherence of communication-based agents and appraisal variables could inform the selection of different communication strategies. Communication-based methods could extend the repertoire of coping strategies. In this article, however, we focus on appraisal-based methods.

4. Emotion and coping

EMA[5] is an implemented process model of appraisal theory following the basic outlines discussed above. EMA embodies a compromise between the theoretical requirements introduced in Section 2 and the pragmatic constraints of building a working "broad agent", including the need to support interactive task-oriented dialogue, real-time control over verbal and non-verbal behavior and responsiveness to external events. Thus, as a theory of appraisal and coping, EMA is by necessity a simplification, albeit one of the most sophisticated simplifications to date, and we discuss its limitations and possible remedies in Section 6 of this article. As a computational model, however, we claim that it is a good level of abstraction for modeling human emotional behavior in an interactive setting, and further, that it is readily

[5] Named in honor of Richard Lazarus (1991) book, *Emotion and Adaptation*, that proposed a unified view of appraisal and coping.

1. Construct and maintain a causal interpretation of ongoing world events in terms of beliefs, desires plans and intentions.
2. Generate multiple appraisal frames that characterize features of the causal interpretation in terms of appraisal variables
3. Map individual appraisal frames into individual instances of emotion
4. Aggregate emotion instances into a current emotional state and overall mood.
5. Adopt a coping strategy in response to the current emotional state

Fig. 3. EMA algorithm.

extensible should additional distinctions be demanded by a particular application.

EMA models appraisal and coping through the five stages listed in Fig. 3 and this section describes each step in detail.[6] For the sake of exposition, we ground this discussion in terms of the following vignette, inspired by Interactive Pedagogical Drama (IPD), an interactive system that teaches coping strategies to parents of children with cancer (Marsella et al., 2000; Marsella, Johnson, & LaBore, 2003). In this example, we model the behavior of an oncologist, Dr. Tom. His patient, Jimmy, is an eleven-year-old boy suffering from stage 4 inoperable cancer. Dr. Tom has exhausted all treatment options and the patient is in extreme pain. The agent interacts with a human participant playing the role of Jimmy's mother. After consulting with a specialist, Dr. Tom concludes the only effective option for controlling Jimmy's pain is to administer large doses of morphine. Dr. Tom opposes this option, however, as it may hasten Jimmy's death.[7] Above all other factors, Dr. Tom values prolonging life, even if the patient is in pain, and especially in someone so young. On the other hand, Jimmy is experiencing intense distress and is fixated on the hope his pain can be reduced. Dr. Tom explains these options to the mother, with the hope that she will decline morphine treatments. If the participant playing Jimmy's mother elects to proceed with the morphine treatments, Dr. Tom feels anger. In the subsequent discussion, we describe how we can represent Dr. Tom's interpretation of this situation, the evolution of his emotional state over time, and the impact of coping strategies.

4.1. Causal interpretation and cognitive operators

The causal interpretation is an explicit representation of an agent's current mental state concerning past, present and future states and actions, their likelihood and desirability, and causal relationships between them. As noted in Section 2.2, we have found it most effective to take classical planning representations as a starting point for our representations, and extend these to incorporate simplified representations of decision-theoretic and intentional information. Planning representations capture a number of distinctions required for deriving appraisal variables, including causal reasoning, the ability to detect future benefits and threats, and the ability to represent the causal agents associated with these benefits and threats. To this we add representations to support decision-theoretic reasoning (i.e, probabilities and utilities) and representations to support reasoning about beliefs and intentions. Here we give a basic overview of the representation, and refer the interested reader to the online reference (Gratch & Marsella, 2004b) for further details.

Fig. 4 illustrates an instantiation of the causal interpretation after Jimmy's mother has authorized the morphine treatments. The interpretation is divided into three parts, the *causal history*,

[6] Note that similar stages have been suggested by other cognitive modeling architectures. For example they are analogous to the standard problem solving cycle proposed by Newell (1990).

[7] Morphine is commonly believed to hasten death by suppressing the respiratory system, though this view is contradicted by some recent studies. What is important is the causal interpretation encodes Dr. Tom's belief that morphine is harmful, not the actual fact of the matter.

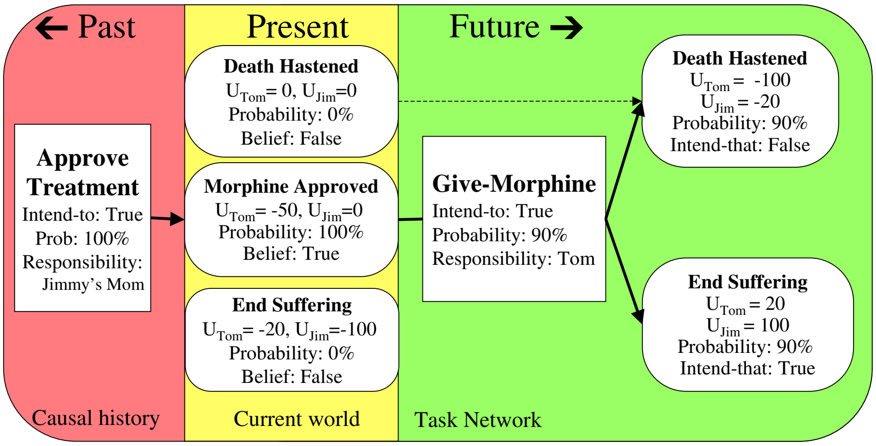

Fig. 4. Dr. Tom's causal interpretation at the end of the scenario.

current world description and the *task network*. The history represents actions that the agent believes (to some degree of certainty) to have occurred in the past, as well as causal relations between actions and states (e.g., the "give morphine" action will cause "death hastened" to be true). The current world description represents an agent's interpretation of the (likely) truth-value of features in the environment (Jimmy's suffering has not ended). The task network encodes future possible plans of the agent or other known entities in the environment (Dr. Tom may administer morphine). This division of the causal interpretation into past, present and future should not be confused with the order in which elements appear in the interpretation. For example, an explanation mechanism might work backward through the causal history, positing more and more remote causes for recent events. Alternatively, the past or future actions might be added as a side effect of natural language understanding: for example, in (Gratch, 2002), agents communicate their future plans to other agents, and the understood sequences of actions are incorporated into the causal interpretation. The division into past, present and future also should not be confused with the certainty of the underlying states and actions. Given the interpretative nature of the causal interpretation, actions in the causal history need not represent what actually occurred but rather an agent's current explanation of what may have occurred, which are can be just as uncertain and subject to revision as future plans.

Specific details of the representation build on prior work on the IPEM planner (Ambros-Ingerson & Steel, 1988), a planning, execution and replanning system that has inspired a number of significant autonomous agent applications (Hill, Chen, Gratch, Rosenbloom, & Tambe, 1997; Rickel & Johnson, 1999). IPEM extends classical artificial intelligence planning techniques by maintaining an evolving current state of the world (rather than just an initial state) and allows actions to have duration and possibly fail. We further extend this system to represent probabilities, utilities, beliefs and intentions, and maintain the explicit history of past actions. The implementation is built atop the Soar cognitive architecture (Newell, 1990).

Actions in the causal interpretation are represented with an augmented STRIPS representation (Fikes & Nilsson, 1971). Actions have an execution state – they may be executed, executing or pending, depending on if they are in the history, current description, or the task network – and are partially ordered. Actions can have duration and effects may occur asynchronously. Causal relationships between actions are represented with establishment relations (causal links) that represent that an effect of an action can or did cause a precondition of another action to be satisfied, and causal threat relations that indicate that some effect threatens or violated an establishment relation. In Fig. 4, the "approve treatment" action establishes the precondition of giving morphine.

The degree-of-belief in the occurrence of an action or the value of a state is characterized by subjective probabilities that are "updated" by inference or observation (i.e., we adopt the "Bayesian viewpoint" of equating probability with degree-of-belief or degree-of-uncertainty). For example, an agent might attribute some a priori probability to the attainment of a goal. This probability can be subsequently refined by generating a specific plan to achieve the goal, essentially re-casting the probability of the goal in terms of the probability of more immediate subgoals. In our example, the a priori probability that Jimmy's pain could be reduced was 95% but was updated to 80% after some planning (see (Gratch & Marsella, 2004b) for implementation details).

Preferences over environmental features are represented as numeric utility distributions over the truth-value of state predicates. Utilities fall in the range of negative to positive one hundred. In Fig. 4, Dr. Tom's opposition to hastening death is represented by a strong negative utility value for this state predicate ($U_{\mathrm{TOM}} = -100$). The causal interpretation also represents the agent's interpretation other agent's preferences. For example, Dr. Tom believes that Jimmy attributes less negative utility to dying (−20). We adopt the common distinction between intrinsic and extrinsic utility: utilities may be either intrinsic (meaning that the agent assigns intrinsic worth to this environmental feature) or extrinsic (meaning that they inherit worth through their probabilistic contribution to an intrinsically valuable state feature). For example, the state "morphine approved" has no intrinsic utility but inherits extrinsic utility of −50 through its probabilistic contribution to the effects of the "give morphine" action (see (Gratch & Marsella, 2004b) for details).

Beliefs are associated with predicates in the current world description. They represent a commitment to the truth-value of a predicate, which may be true, false, or unknown. The perception module deliberately commits to one of these truth-values based on evidence (such as the probability that the predicate is true). Following Grosz and Kraus (1996), EMA uses "intend-to" and "intend-that" to signify the attribution of intention over actions and states. Note that the causal interpretation currently embodies a much simpler logic of belief and intention than Grosz and Kraus (1996), ignoring the issue of mutual belief or joint intention (though see (Traum, Rickel, Gratch, & Marsella, 2003a)).[8]

The causal interpretation defines an agent's current mental state but an agent must also possess some mechanism for updating mental state based on input from the outside world or through deliberation. In EMA, these mechanisms are divided into a set of fine-grained operators, motivated by Newell's (1990) Unified Theory of Cognition. These may be viewed as individual procedures that compete for attention in parallel but are selected and executed serially. In Newell's theory, these are posited to operator on the 100ms scale in human cognition. In our current implementation, operators include planning related operators (e.g., add a plan step, update a belief, update an intention, etc.), dialogue related operators (e.g., understand speech, output speech, update dialogue state), and execution and monitoring operators (e.g., monitor an effect, action initiation, etc.). The set of these operators and their details are unimportant in the context of this paper, though we will make reference to them in the section on emotional focus. Collectively we refer to these as *cognitive operators.*

4.2. Appraisal frames and variables

The causal interpretation encodes many past and future events that may be of significance to the agent. EMA builds on our earlier work on the

[8] Note that our model distinguishes between assigning high utility to a state and intending that a state be achieved. The former is a statement of preference, the latter a statement of intent. An action is only intended if the agent intends that some state be achieved, the action is relevant to achieving that state and the agent deliberately adopts the intention to perform the action. An agent may intend an outcome that it considers undesirable, for example, to satisfy a social obligation or imperative. In the example, Dr. Tom intends to perform an action (give morphine), whose outcome (hasten death) he does not desire.

Émile system to appraise the significance of events (Gratch, 2000). Each of these assessments is represented by an *appraisal frame* that contains the derived value of appraisal variables associated with the event from a given perspective. As the agent's causal interpretation changes as the result of cognitive processes, new appraisals are formed and old ones are retracted. EMA appraises multiple event simultaneously and each event from multiple perspectives (its own perspective as well as the imagined perspective of other).

A "significant event" is defined to be any action in the causal interpretation that has an effect that facilitates or inhibits a state predicate with non-zero utility. By facilitation, we mean there is an establishment relation between the action and the state, whereas inhibition means there is a causal threat relation between the action and the state. Under this definition, any single external action may result in multiple appraisal frames: actions may have multiple effects and any single effect may facilitate or inhibit multiple states. These states may be of perceived interest to multiple individuals. EMA creates a separate appraisal frame for each facilitation or inhibition relation from each perspective. These frames are not necessarily created at the same time, but are only triggered when reasoning modules represent the facilitation, inhibition, or preference in the causal interpretation. Thus, a single action may engender multiple and possibly conflicting emotional responses.

In realizing the theoretical framework, we have focused on a subset of the appraisal variables listed in Table 1. Specifically, we model relevance, desirability, likelihood, causal attribution, and coping potential, although, with the exception of ego involvement, we believe the other appraisal variables listed in Table 1 could be straightforwardly added to the model (these limitations are discussed further in the conclusion of the article). The detailed appraisal rules used by EMA are listed in (Gratch & Marsella, 2004b). Here we describe their behavior at a high level.

4.2.1. Perspective

Though not explicitly treated as a variable by appraisal theories, a key aspect of appraisal is the ability to reason about how events or one's own actions impact other individuals. Given that the causal interpretation corresponds to the agent's own subjective interpretations, any assessment of the agent–environment relationship is ultimately from the agent's own perspective. However, it is clear that people consider the emotional impact of events on other entities preference structures quite different from their own. For example, the joy at winning a game may be mitigated by the imagined distress felt by one's opponent, even if this distress is not visibly displayed. Such reasoning from the perspective of others seems to lie at the heart of the more social emotions such as guilt or shame.

Following Elliott (1992), we model this notion of perspective explicitly. For a given event, EMA constructs multiple appraisal frames, one from the perspective of the appraising agent, and one from each "imagined" perspective of other agents that the appraiser is modeling. In theory, each appraisal variable associated with an event could differ in value base on the perspective. For example, Dr. Tom might believe that Jimmy's death is likely and undesirable while Jimmy, given his pain, may view death as desirable though unlikely. All of the variables described below are treated as conditional on an agent's perspective; however our current implementation only supports differences in preferences agents have over events. See Section 6 for a discussion on how to relax this limitation.

4.2.2. Relevance

Relevance measures the significance of an event for the agent. In EMA we equate significance with utility. An event outcome is only deemed significant if it facilitates or inhibits a state predicate with non-zero utility. In practice, virtually every event will have at least some indirect impact on intrinsically important states, and thus almost every event will have some relevance. In the interest of efficiency, EMA only constructs appraisal frames for event outcomes where the magnitude of the utility, from some agent's perspective, exceeds some small fixed threshold (which we set at 1.0 in our applications).

4.2.3. Desirability

Desirability captures the appraised valence of an event with regard to an agent's preferences. An event is desirable, from some agent's perspective, if it facilitates a state to which the agent attributes positive utility or if it inhibits a state with negative utility. An event is undesirable if it facilitates a state with negative utility or if it inhibits a state with positive utility. Desirability does double duty in appraisal theories of emotion, acting to categorically separate certain emotions (Joy from Distress; Hope from Fear) and acting as an intensity variable (i.e., a factor that moderates the intensity of the response). In EMA, the magnitude of the positive or negative utility serves as an intensity variable. Events are not appraised from an agent's perspective if the consequence has no utility for the agent (such events are treated as irrelevant from the perspective of appraisal).

4.2.4. Likelihood

Likelihood characterizes the certainty of the event. Is it something that definitely happened or definitely will happen, or is it something that may have or might occur. Likelihood tends to do double duty in appraisal theories, acting to categorically separate certain emotions (Joy from Hope; Distress from Fear) and acting as an intensity variable. EMA equates likelihood with the probability of an event. We define likelihood as a threshold over the probability of an event to categorically state if an event may be either certain or uncertain.

4.2.5. Causal attribution

Causal attribution characterizes whether the causal agent behind an event deserves credit or blame. In general, the problem of attributing credit or blame involves a number of considerations (Shaver, 1985; Weiner, 1986). Who caused the outcome? Did they foresee the consequence? Was it intended? Where they coerced? Ideally, the attribution would account for all of these factors. In our current implementation, causal attribution is assigned to whatever agent actually executed the action in question. Whether an agent deserves credit or praise depends on whether the outcome is desirable from the perspective being taken. The weight of the praise/credit depends both on the desirability of the corresponding event and its likelihood. Note that some psychological theories define praise more narrowly, in terms of the upholding or violation of social norms, notably (Ortony et al., 1988). Our implementation is less constrained. Norms can be encoded as preferences over the outcomes of actions (one can distinguish the means from the ends by characterizing different means by more or less desirable side effects) and agent can be praised or blamed for causing non-normative benefits or harms.

4.2.6. Controllability

Two appraisal variables, controllability and changeability, are associated with an agent's ability to cope (in a problem-focused way) with appraised events. Controllability is a measure of an agent's potential to actively reverse negative, maintain positive circumstances. It is computed by identifying any actions in the causal interpretation that have effects that impinge on the appraised event. For example, if a desired goal is threatened by an action and some other act in the current causal interpretation could re-establish the goal – i.e., a white knight in planning terminology (Chapman, 1987) – then the causal agent associated with this action is presumed to have a measure of control equivalent to the likelihood of this re-establishment. Alternatively, if an undesired state may (has) come to pass, any actions in the causal interpretation that undo the state are examined. Currently, we base controllability on the maximum of the likelihood of all such causal influences.

4.2.7. Changeability

Changeability is a measure of how likely an appraised event will change without direct intervention by some agent. On our model, this corresponds to the likelihood that the event will change by some factor other than direct action by the agent from whose perspective the frame is being assessed. For example, a looming threat may be appraised as changeable if the effect of the threatening action is uncertain or if some intervening act not under control of the agent might occur.

4.3. Emotion instances

EMA maps each appraisal frame into an instance of emotion of a specific type and intensity,

Table 3
Emotion categorization and intensity rules

Appraisal configuration	Emotion	Intensity
Desirability(p) > 0, Likelihood(p) < 1.0	Hope	Desirability(p) × Likelihood(p)
Desirability(p) > 0, Likelihood(p) = 1.0	Joy	Desirability(p) × Likelihood(p)
Desirability(p) < 0, Likelihood(p) < 1.0	Fear	$\lvert$Desirability(p) × Likelihood(p)$\rvert$
Desirability(p) < 0, Likelihood(p) = 1.0	Distress	$\lvert$Desirability(p) × Likelihood(p)$\rvert$
Desirability(p) < 0, causal attribution(q) = blameworthy	Anger	$\lvert$Desirability(p) × Likelihood(p)$\rvert$
Desirability($q \neq p$) < 0, causal attribution(p) = blameworthy, causal agent = p	Guilt	$\lvert$Desirability(q) × Likelihood(p)$\rvert$

which are subsequently focused and aggregated. In our current implementation, EMA uses a simplification of the mapping proposed by Clark Elliott (1992), which in turn inspired by Ortony et al. (1988).[9] Elliott's mapping supported 24 different emotion categories types. Here we illustrate six (Hope, Joy, Fear, Distress, Anger and Guilt), which have sufficed for our current applications.

Table 3 lists the basic rules that map from the configuration of variables in an appraisal frame to an emotion instance of a particular category and intensity. Note that each appraisal variable is conditional on some perspective, p. Hope arises from a belief that something good may happen or have happened. In EMA, this translates into the fact that desirable future state may be achieved or an undesirable future state may be averted. Currently, EMA assumes that past events and outcomes are known with certainty so hope can only arise from future eventualities. This restriction could be relaxed if EMA were to incorporate some form of retrospective inference, thereby allowing hope to arise from uncertain past attributions as well (e.g., I hope my friend wasn't in the World Trade Center that day). The intensity of hope is based on the state's desirability and its likelihood of achievement. Joy arises when something desirable has happened or is seen as inevitable (occurs with probability one). This translates into the fact that a preferred state has been attained where intensity is tied to the state's desirability. Fear arises from a belief that something bad may happen or has happened. In EMA, this means some goal is unestablished, or its establisher is threatened. As with hope, fear is currently restricted to future events. Intensity is based on the state's desirability and its likelihood of failure. Distress arises when some fear has been confirmed. This translates into the fact that a desired state has been prevented from occurring (its establishing plan was threatened and the threat occurred), or an undesirable state has been achieved or is deemed inevitable. The intensity of distress is directly proportional to the states desirability. Anger arises when some agent is responsible for (possibly) producing an undesirable state (note that under our model, anger may be self-directed). Guilt arises when an agent is deemed blameworthy for (possibly) causing an outcome that some other agent, q, is believed to find undesirable.

4.4. Emotion focus and aggregation

Clearly, we are awash in potential appraisals, stemming from our memories, our daily experiences and events in the larger world and our expectations about the future. Our computational approach to appraisal acknowledges this fact by maintaining numerous simultaneous appraisals that are updated by changes to the causal interpretation: the causal interpretation encodes acts in the past, present and future; a single act may elicit multiple appraised outcomes; and each outcome may be appraised from multiple perspectives. But

[9] Appraisal theories map configurations of appraisal variables into emotions of some class and intensity, though theories differ on the form of this mapping and the number of distinguishable classes. For example, many theories argue that the distinction between fear and anger depends on the appraised sense of control (Lerner & Keltner, 2000). As our focus is on explicating underlying mechanism rather than defining specific emotion types, we are agnostic on this mapping and other mappings are easily supported.

given that our virtual human could be awash in such appraisals, what focuses it on particular appraisals?

The psychological literature suggests that this focus is influenced indirectly by the organism's emotions, moods and coping strategies. For example, a fearful individual will be biased towards pessimistic interpretations of subsequent events (Lerner & Keltner, 2000). The need for a model of focus is also suggested by a range of coping strategies that attempt to divert attention from stressful events. For example, people cope by avoiding difficult decisions or by undertaking other activities as a distraction (Lazarus, 1991). This suggests that the coupling between stressful events and emotions is mediated by some form of attention to those events. EMA models such influences on appraisal through a generate and biased-test approach. The model generates multiple potential appraisals and then narrows this set through an interaction of cognition, emotion and mood, as discussed below.

4.4.1. Emotional focus

We adopt an attentional model motivated by several psychological theories (Frijda & Zeelenberg, 2001; Smith & Kirby, 2001) to compute a moment-to-moment subset of appraisals to bring into focus. The key idea is to tie emotional focus to the agent's cognitive operators. As discussed in Section 4.1, our model posits a number of atomic operators that vie for access to the causal interpretation to retrieve information or post intermediate results (e.g., interpreting a speech act, updating a belief, etc.). Whenever a cognitive operator accesses a portion of the causal interpretation (e.g., states or actions), any appraisal frames associated with those data structure are brought into focus. For instance, in our doctor agent example, a question from another agent or user, such as "What are you going to do for the cancer patient?" requires the natural language understanding module to access the causal interpretation to find the referent for "cancer patient" as well as the referent for actions that impact him. This would bring into focus emotions associated with Jimmy and the impending use of morphine. Note this focus mechanism is similar to how spreading activation works in cognitive models like Act-R, whereby a concept is brought into working memory by some task and activation is subsequently spread to related concepts in long-term memory (Anderson, 1993).

This approach to focus is notable both in its simplicity and its explanatory power. For example this coupling of cognitive operations and appraisal/coping ensures that not only are emotions guided by cognition but also that emotions and coping can inform the cognitive operations as well. For example, the appraisal and coping mechanisms help in disambiguating dialog, by indicating which alternative interpretation of an ambiguous speech act is associated with the most intense appraisals, and therefore is the most salient interpretation. Thus the doctor agent would interpret the previous question as being about the issue of relieving the suffering with morphine.

In addition, this focus mechanism provides an elegant explanation why various distraction-based coping operations such as disengagement work. Certain coping strategies work by making portions of the cognitive interpretation less accessible to cognitive operations, and therefore making the associated appraisals less likely to come into focus. For example, by dropping an intention, the planner is less likely to access the state or task that the intention was associated with. This mechanism could also support more subtle strategies like going to a party to distract oneself from thinking about a stressful term paper. Ensuring the cognitive operations associated with writing the paper will not come into focus puts that stress out of mind, at least temporarily.

As we have realized this focus mechanism, it implements a rather conservative, very rational, approach to relating appraisal to coping. The emotion the agent feels concerning some event guides how the agent copes with that event. However, in human behavior, the relation between emotion and coping response is not always so clear. For example, some event at work may anger a person and that anger may bias how the person copes with some unrelated event later in the day at home. In such behavior, there is not such a clear connection between appraised event and response here. Rather the emotion seems to persist and affect later behavior. We will come back to this issue in the discussion section, but a person's mood likely plays an important role.

4.4.2. Mood

Moods are an affective phenomena closely related to emotion. Typically, mood is distinguished from emotion as being more global, diffuse and longer-lasting. Moods are not "clearly related to a single object or piece of business in an adaptational encounter, as is the case in acute anger or fear" (Lazarus, 1991). Moods are important to model because they have been shown to impact a range of cognitive, perceptual and behavioral processes, such as memory recall (mood-congruent recall), learning, psychological disorders (depression) and decision-making.

EMA maintains an aggregate emotional state that corresponds to the agent's mood. In contrast to the focused emotions, mood is computed by aggregating every emotion instance associated with the current causal interpretation. For each emotion type (e.g., Joy, Fear, etc.), EMA simply adds the intensities of all elicitors of that type. The aggregate values are then passed through a sigmoid function to map the emotional state to a value from zero to one. The emotions identified by the focus mechanism are shorter-term as they are tied to specific cognitive operators, whereas mood is tied to the overall causal interpretation, and thus changes more slowly. Note one might model mood in other ways. Rather than an aggregation of emotional states it could be modeled as a byproduct of underlying physiological or biological processes that in turn influence appraisal and coping processes. Another alternative is to model mood as the product of appraising "larger, longer lasting, existential issues about the person's life and how it is going". We will return to this in Section 6.7.

Mood is used in combination with the focus mechanism to determine a single in-focus emotional instance. The focus mechanism identifies several emotional instances associated the current cognitive operator. Currently, EMA adds the current mood to the intensity of each of these instances and uses these 'mood-biased' intensities when determining the most intense emotion to place into focus. For example, if an event leads to equally intense appraisals of hope and fear, the agent will focus on one aspect or the other depending on its congruence to the overall mood. With such a mechanism, the mood can bias any cognitive processes influenced by emotion, including coping.

4.5. Coping

Our research on a computational model of coping is ongoing (Marsella & Gratch, 2002, 2003). Here we characterize the current state of that research. The challenge in our coping work is to translate coping strategies, like those presented in Table 2, into concrete guidance for future action or concrete changes in how the agent views its relationship with the environment. This challenge is made more difficult because the psychological literature defines coping strategies in a somewhat nebulous fashion. Nevertheless, we argue that coping strategies can be defined in terms of the same representational features that underlie appraisals.

Our approach to coping tightly couples the process that leads to emotion, appraisal, with the coping process that deals with them. In essence, coping is cast as the inverse of appraisal. To discharge an emotion about some situation, one obvious strategy is to change one or more of the factors that contributed to the emotion. Coping operates on the antecedents of appraisals – beliefs, goals and plans – but in reverse, seeking to make a change, directly or indirectly, that would have the desired impact on appraisal. Coping could impact the agent's beliefs about the situation, such as the importance of a threatened goal, the likelihood of the threat, responsibility for the threat, etc. Further, the agent might form intentions to change external factors, for example, by performing some action that removes the threat. Indeed, our coping strategies, can involve a combination of such approaches. This mirrors how coping processes are understood to operate in human behavior whereby people may employ a mix of problem-focused coping and emotion-focused coping to deal with stress.

To computationally model this process, we use the causal interpretation as the common representation used by both appraisal and coping. Appraisals are driven off features of the causal interpretation. Coping strategies act by altering features. Because it operates over the causal interpretation, coping's impact may be either immediate (abandoning a goal will alleviate stress arising from a blocked goal) or indirect (as when a changed preference alters future planning behavior). In addition, the strategies may impact beliefs about the past, present or future as well as current and future intentions.

In modeling this process, we have generalized somewhat what is meant by coping. Coping is often construed as a response to strong negative emotions; however, we view it as a general response to all kinds of emotions, strong and weak, negative and positive. This view is supported by a careful consideration of the coping strategies. Strategies such as active problem solving, wishful thinking, seeking social support and suppression of competing activities are just as applicable to achieving a desire as to addressing a threat. For example, a child desiring a toy may engage in all the above strategies: getting a job after school to purchase the toy (planful problem solving), wish that some relative would give it to him (wishful thinking), ask his parents to buy it for him (seek social support), drop out of after-school activities so he could earn more money to purchase it (suppression of competing activities).

4.5.1. Coping process

To realize this model, we propose a concrete mapping between commonly identified coping strategies and representational features of the causal interpretation. In laying out this mapping we must address several issues. Given that there are multiple appraisals, which appraisals lead to coping? What is the specific mapping from a strategy to representational features? If a strategy has multiple instantiations or multiple strategies apply, how do we arbitrate between strategies? Specifically, what situational factors may mediate which strategy is selected? We address these issues within a five-stage process: (1) identify a coping opportunity, (2) elaborate coping situation, (3) propose alternative coping strategies, (4) assess coping potential, and (5) select a strategy to apply.

4.5.1.1. Identify coping opportunity. Whenever the agent performs a cognitive operation, for example, updating an intention or understanding speech, coping identifies any associated appraisal that could motivate coping. To do this, coping creates a *coping elicitation frame* that consists of a number of coping related fields.

The *focus-agency* is the agent or object that "provoked" the cognitive operation (for example the speaker in the case of understand speech or the agent itself in the case of planning operations).

The *interpretation-objects* are any tasks, states or individuals in the causal interpretation referenced by the cognitive operation. There may be multiple referents. For example, if a speaker asks "what happened", the referents could be any task in the causal history. For each interpretation object, coping identifies the strongest positive and negative appraisals associated with the referent. For example, if the "give-morphine" task is the referent, the appraisals associated with hastening death and reducing suffering would be the most negative and positive appraisals, respectively. Coping also identifies an *agency-max*, which corresponds to the max emotion that the agency believes the focus-agency has about the same referent.

The *max-interpretation* is the interpretation object with the strongest appraisal. If the intensity of the max appraisal of the max-interpretation exceeds some pre-specified constant, the coping elicitation frame is identified as a coping opportunity.

4.5.1.2. Elaborate coping situation. The elicitation frame is also elaborated with various situational factors that impact the selection of an appropriate coping strategy. Currently, this includes the social relations between the various individuals identified in the interpretation objects and focus agency as well as their actual or potential responsibilities with respect to the emotionally significant event.

4.5.1.3. Propose alternative coping strategies. Coping strategies are proposed for each coping opportunity based on features of the coping elicitation frame. Each strategy consists of two parts, a set of conditions that define its applicability, and an abstract characterization of its effect on the causal interpretation. We will detail the strategies later in this document, but as a quick example, a problem directed strategy might have as its applicability conditions that the coping frame most intense appraisal be a threat to a desired goal (e.g., giving morphine hastens death). The effect of this strategy is that some change must be identified that overcomes this specific threat.

4.5.1.4. Assess coping potential. The assessment of coping potential takes a strategy's abstract effect and maps it into one or more elements of the causal interpretation that, if changed, would alter the appraisals in a desired way. There may be multiple ways to achieve this direction and the assessment of potential also ranks these alternatives in terms of their expected impact on the appraisal frame. For example, a problem directed strategy to address the threat caused by giving morphine might address the threat either by identifying one or more tasks that could reverse the undesired effect of giving morphine (adding a "white knight") or by dropping the intention to give morphine. In the case of planful strategies, these assessment rules correspond to fairly standard plan critics (e.g., find some task that possibly confronts a precondition of a threatening task).

4.5.1.5. Select one strategy. Finally, coping picks one or more strategies and applies them. If there are multiple applicable strategies, the application process currently works sequentially. Heuristics establishes preference over the strategies based on an estimation of the ability to control a situation and the likelihood that it will get better on its own. The preferred strategy is applied and if there is additional strategies that are still applicable after that application, they are in turn applied in order of preference.

4.5.2. Coping strategies

We now will consider in greater detail the coping strategies we have modeled. Several of the strategies listed in Table 2 have been implemented.

4.5.2.1. Planning. Planful coping involves forming an intention to take an action whose effect achieves the desired state or blocks direct or indirect threats to the desired state. If the max appraisal associated with a coping elicitation frame is positive (e.g., a desirable state was achieved or may be achieved in the future), the strategy asserts a preference to maintain this state. Similarly, if the max appraisal associated with the coping frame is negative (e.g., a desirable state was threatened) the strategy identifies actions that would overturn the threatening circumstances. During the assessment of coping potential, plan critics fire, attempting to identify specific tasks that, if they were augmented with positive or negative intentions, would have the desired effect. For example, if the doctor feels good about reducing suffering, he might form an intention to give morphine. The plan critics that assess coping potential correspond to conventional plan critics (Chapman, 1987) – e.g., if a step clobbers a desired step P, considering adding a step that reestablishes P (a white knight).

Planful strategies impact appraisals indirectly by motivating future planning. For example, if coping forms an intention to perform a task, the planner will be invoked to attempt to achieve the preconditions of that action. As this will change the causal interpretation it may lead to new appraisals and subsequent coping.

4.5.2.2. Positive reinterpretation. Positive reinterpretation involves finding positive meaning in some otherwise negative event. Computationally, this means finding some direct or indirect consequence of the event that is desirable and emphasizing it by increasing its utility for the agent. For example, giving morphine has the negative consequence of hastening death but at least it reduces suffering. During the assessment of coping potential, rules identify any immediate consequences with positive utility, or any consequences that are facilitated indirectly via intermediate causally connected tasks. Currently, we allow utility values to be incrementally adjusted within a user-specified range and if adjustment is possible, these consequences become candidates for change. If adopted, the utility of one of these candidates is adjusted upward.

Positive reinterpretation will lead negative events to be re-appraised in a more positive light. This may lead indirectly to the formation of new intentions. For example, the doctor may initially not intend to give morphine because on balance he believes its cost exceed the benefits. Following positive reinterpretation, he may believe that the benefit now exceeds the cost and give the drug.

4.5.2.3. Acceptance. Acceptance is the recognition that a negative appraisal is unavoidable. Computationally, this corresponds to the situation where the maximum appraisal is a threat to a desirable intended state. Under these circumstances, this strategy proposes dropping the intention, essentially dropping the commitment to achieve this state.

Acceptance will lead the planner to stop the search for plans to achieve the desired state. So while the threat will still be appraised as undesirable, through the focus of attention mechanism, the undesirable appraisal should come into focus less often as cognitive operations such as update-intention and update-belief will no longer reference the state. For example, if the doctor accepts that hastening death is unavoidable he may become less focused on that consequence and be more inclined to provide morphine.

4.5.2.4. Denial/wishful thinking. Denial works by denying the reality of an event. The strategy is proposed if the most intense appraisal associated with the coping frame is negative. During the assessment of coping potential, rules identify factors leading to the negative appraisal that are candidates for denial. If selected, one of these candidates is manipulated to appear less likely. For example, one way to mitigate the distress associated with providing morphine is to deny to oneself that morphine hastens death. The strategy adjusts downward the probability that an effect of an action will occur, where the adjustment falls within some user-specified range.

The consequence of denial is that certain threats or establishment relations will appear less likely. This will directly reduce the intensity of the negative appraisal. This may also indirectly impact planning and plan execution behavior. For example, the planner may not confront certain threats if they appear, through denial, to be unlikely.

4.5.2.5. Mental disengagement. Mental disengagement acts by reducing an agents "investment" in some state of affairs. Computationally, this corresponds to a character lowering the assessed utility of a previously desired state. For example, if the doctor is distressed about give morphine, he may distance himself from the situation by lowering the utility of all of states associated with the action. This is different than acceptance where the agent drops the intention but still maintains its preference for the outcome.

Mental disengagement lowers the emotional charge associated with the event. It may also lead the agent to indirectly drop intentions associated with the event as the overall desirability of the associated actions are reduced.

4.5.2.6. Shift blame. People also employ various coping strategies that revolve around manipulating blame, specifically self-blame and other-blame. For example, a person may shift blame to someone else. The doctor could decide that Jimmy's mother has taken responsibility for the act of giving-morphine, in which case he would subsequently appraise anger towards toward her for the, from his perspective, undesirable consequences.

4.5.3. Mixed coping strategies and consistency

Since our strategies work on a set of representational features, they can potentially operate in tandem as long as they are consistent in terms of the proposed changes to the causal interpretation. The doctor may behaviorally disengage from ending suffering by dropping the intend-to give-morphine while simultaneously engage in wishful thinking that the suffering will be less probable or that some fortuitous event will intercede to reduce it. This tandem, combined operation is feasible as long as the various strategies don't conflict in their manipulations of the causal interpretations. Alternatively, the Doctor may become resigned to the fact that death is inevitable and therefore not caused by the morphine. Specifically, he could reduce the utility of inhibiting that state or drop the intend-that while simultaneously find positive reinterpretation in ending the suffering by increasing the utility of that goal. By allowing coping strategies to combine, a few simple strategies can realize more complex coping behavior. Further, this behavior can be allowed to unfold over time, as consistent strategies are applied in turn.

This question of consistent changes raises an interesting challenge for intelligent agent design. Coping is making changes to beliefs about likelihood and responsibility, changes to desirability, forming wishful intentions, etc. Though psychologically plausible, this is clearly unorthodox from a traditional logical or decision-theoretic interpretation of these terms. One can view coping as an alternative, psychologically motivated calculus for updating subjective probabilities and utilities. But as we have presented it, this calculus is clearly constrained. An agent shouldn't be free to simply wish away important goals or beliefs. Our current approach to this problem is to make incremental changes. So, for example, the likelihood of a wished-for event only changes a small increment in the direction indicated by the strategy. If the same coping strategy is selected again and no other observation or aspect of the causal interpretation is in conflict, the utility or probability is further incremented. On the other hand, if the world intervenes and "sets the agent straight," the changes are reversed.[10]

Although our current approach is far from a complete solution, it is nevertheless interesting because it raises the issue of how certain coping strategies interact. For example, consider a person that has coped by altering a belief. Subsequent coping strategies can serve to protect that belief change. A person might "avoid" social interaction to avoid opposing views towards that belief, or alternatively, a person might "seek social support" to get confirmation of the new belief. (Note this suggests belief maintenance as a type of maintenance goal.)

Overall, coping can be viewed as a generalization of problem solving that encompasses not only the traditional AI view but also meta-level control of problem solving and belief revision. For example, as Dr. Tom's goals are threatened and emotions arise, coping strategies in essence move problem solving to a meta-space where alternative approaches to addressing the problem are considered. New goals may be adopted, old goals dropped, the value of achieving a goal may be reassessed in a number of ways, beliefs about the goal or threats to it may be altered, etc. Essentially, these strategies can be viewed as the individual re-fitting his behavior to the environment.

4.6. Process example

We illustrate the dynamics of EMA by walking through the medical example. In the beginning of this example, Jimmy is in pain ("End Suffering" is false), his death has not been hastened ("Death Hastened" is false) and morphine has not been approved ("Morphine Approved" is false). There is no causal memory of how these states occurred (they are established by an "*init*" step that encodes the initial situation as its effects). As the situation unfolds, Jimmy's mother (the human participant) asks Dr. Tom (a virtual human) to reduce her son's pain. This is interpreted as a request to achieve the goal "End Suffering" (implicitly referring to Jimmy's suffering), represented by a state in the task network (see Fig. 5). Dr. Tom prefers that patients not suffer, represented by a utility distribution over the truth values of the "End Suffering" predicate: "End Suffering" has modest negative utility when it is false and modest positive utility when it is true (utility values are drawn from a range [−100..100]). Dr. Tom further believes that Jimmy assigns high negative and positive utility to the truth values of this same state. Where these utilities come from is outside the scope of the present work. Here, they are hard coded constants in the domain theory we provided to EMA. One might imagine more sophisticated models of attributing preferences. For example, if Dr. Tom has an empathetic personality, he might adopt preferences influenced by those around him. Finally, Dr. Tom believes a priori that this goal is relatively easy to achieve.

[10] EMA's implements a local hill climbing solution. This is in contrast to the global coherence approach of Thagard (2000) wherein a belief change is feasible to the extent it globally coheres with other related beliefs and goals. It may be possible to model such coherency by considering the overall change in emotional intensity that would arise from an altered belief, rather than focusing only on the max appraisal, though this requires extending the causal interpretation to support counterfactual reasoning: e.g., how would I feel if X were true.

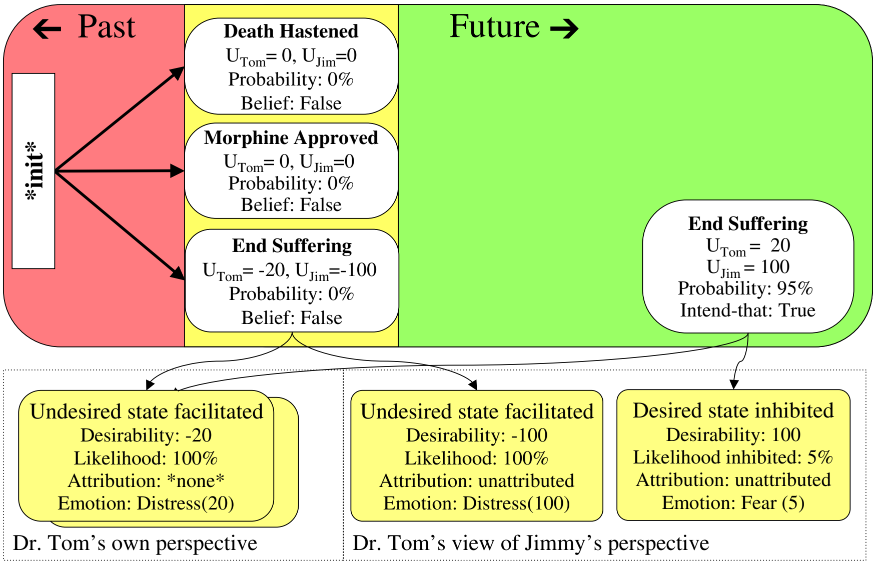

Fig. 5. Dr. Tom's causal interpretation after Jimmy's mother's request to end his suffering (represented as a desired and intended state in the future). Dr. Tom attributes different preferences for this goal to Jimmy.

Appraisal rules fire as a consequence of the change to the causal interpretation and assess it from multiple perspectives. Fig. 5 illustrates how this initial causal interpretation is appraised from Dr. Tom's own perspective and Jimmy's imagined perspective. There are events in the interpretation with non-zero utility: the facilitation of NOT (end-suffering) by the initial step; and the inhibition of the goal to end-suffering, which is currently blocked by the lack of a plan to achieve it. The negation of "End Suffering" is undesirable to both Dr. Tom and Jimmy, leading two appraisal frames which both characterize it as undesirable, confirmed and with un-attributed credit. Each frame generates an elicitor of distress. In Jimmy's case, the distress is imagined to be quite strong ($|\text{Desirability(Jimmy)} \times \text{Likelihood(Jimmy)}| = 100$). In Dr. Tom's case, the appraised distress is moderate. Similarly, the future state of end suffering is appraised from both perspectives, leading to fear rather than distress as the outcome is uncertain. From both perspectives the fear is small as there is only a 5% chance that a plan will not be found. Appraisals are treated differently depending on whose perspective they are appraised from. Emotions appraised from the Dr. Tom's own perspective are folded together and directly influence the agent's overall emotional state, influencing facial expressions, body language, speech and decision-making. Emotions appraised from Jimmy's perspective do not directly influence Dr. Tom's emotional state.

Given that reducing Jimmy's pain has positive utility and seems easy to achieve, Dr. Tom is motivated to act on this goal. The agent agrees to the request, responding with an affirmative speech act, and thereby obligating its self to achieving the future state. This obligation is represented by an intend-that attribute in the causal interpretation. The planning module, seeing an unsatisfied but intended goal, searches for a plan. The planning algorithm finds only one way to achieve this goal – the single-step plan to give Jimmy morphine – and the a priori probability that Jimmy's suffering will end is updated according to this new information. The new action, its preconditions, and an undesired side-effect (hasten death) are added to the causal interpretation (Fig. 6). As this is the only way to achieve an intended goal, the action and its preconditions are marked as intended via the standard axioms of intention. The precondition of this action (morphine approved) has negative extrinsic utility for Dr. Tom, inferred by backing up

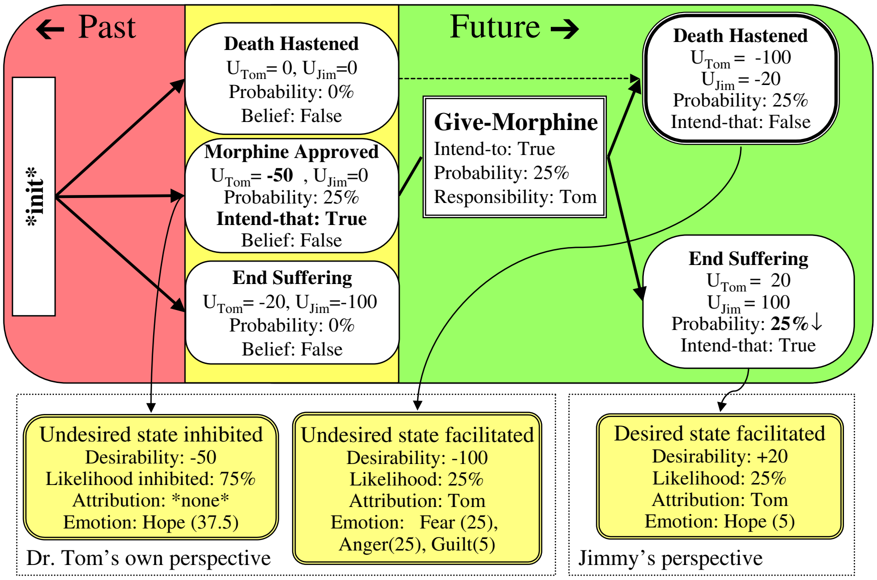

Fig. 6. Dr. Tom's causal interpretation after some initial planning. The "Give-Morphine" task was added with its precondition and effects. This influences the likelihood that Jimmy's pain will end. (Differences from the previous figure are highlighted with double lines or in bold and changes in values are noted with arrows.)

the utilities of the action's effects through to its precondition. A priori, this precondition has low probability of achievement, reflecting that Dr. Tom doesn't believe that Jimmy's mother will approve of the treatment.

These changes to the causal interpretation trigger changes to the existing appraisal frames and the formation of additional frames. Fig. 6 highlights a few key appraisal frames. As the probability of goal attainment is now reduced, the hope associated with Jimmy's suffering is reduced. In addition, new frames appraise the expectation that the morphine will hasten Jimmy's death, and that the subgoal of giving treatment is unachieved. Dr. Tom experiences a small amount of fear associated with hastening death (small because the probability of goal attainment is small). The agent is more hopeful that the use of morphine will not be authorized. This is an interesting case because it illustrates how, in our model, an agent can work towards a plan that it is loath to achieve. The agent intends that morphine be approved – propagated down from the intention of ending suffering – but the agent would prefer this state remains unsatisfied (it has negative extrinsic utility). Thus, the agent is hopeful its plans will fail, and this in turn can motivate dialogue strategies where Dr. Tom tries to convince Jimmy's mother to withhold authorization (Traum et al., 2003a).

Dr. Tom's hope is short lived. Jimmy's mother asserts that she approves of the treatment. This assertion is represented in the causal history by an "approve treatment" action that establishes the precondition. The mother is noted as being responsible for this act (see Fig. 7). Nothing prevents the treatment from going forward and this dramatically increases the probability that the consequences of giving morphine will be achieved (each effect has a 90% achievement probability). With these changes, the appraisal frame underlying Dr. Tom's hope retracts, replaced with an appraisal leading to distress. As Jimmy's mother is deemed blameworthy for this unfortunate consequence (she caused an outcome with negative utility for Dr. Tom), EMA generates an instance of anger.

This negative emotionality can motivate the agent to adopt one or a number of coping strategies. Fig. 8 illustrates the causal interpretation after Dr. Tom has adopted the strategy of denial. In this case, the strategy was applied to the undesirable outcome of giving morphing (e.g., hastening

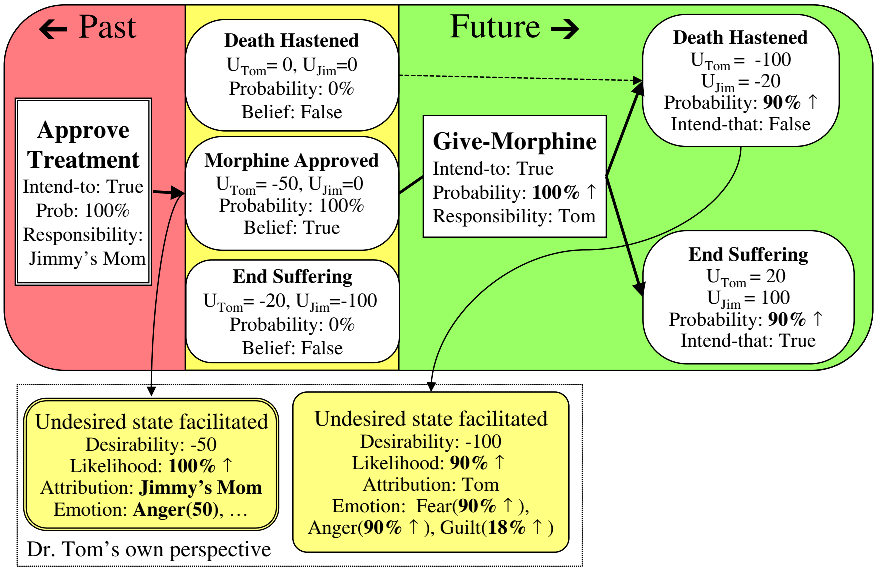

Fig. 7. Dr. Tom's causal interpretation after learning the mother approves of the use of morphine. This raises the likelihood that both effects of "Give Morphine" will be achieved. Note that the status of "Morphine Approved" has shifted from being inhibited (by the lack of an establishing action) to being facilitated.

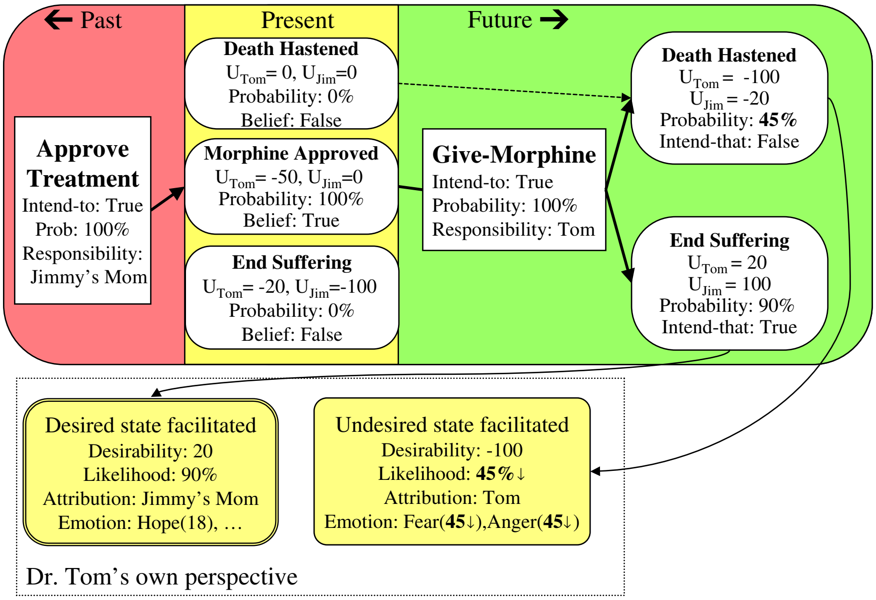

Fig. 8. Dr. Tom's causal interpretation after engaging in denial. The perceived likelihood of "Death Hastened" is lower. After reappraisal, the negative emotions resulting from this outcome are less intense.

death). The consequence of the strategy is to reduce the perceived likelihood of this negative outcome ("well, sometimes morphine doesn't hasten death; perhaps Jimmy will be one of those cases"). As a result, the intensity of Dr. Tom's fear and anger drop.

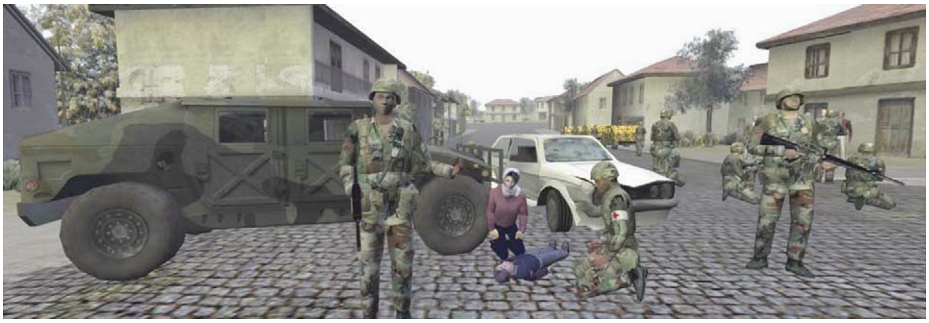

Fig. 9. A scene from the Mission Rehearsal Exercise.

5. Application

One contribution of our work is that we can demonstrate a computationally coherent account of emotion processing, however it remains to show that this model can practically inform real-world applications. Here we briefly describe how the work contributes to the design of a virtual reality training environment that teaches decision-making skills in high-stakes social situations (Gratch & Marsella, 2001; Marsella & Gratch, 2001; Rickel et al., 2002). Within this Mission Rehearsal Exercise (MRE) training system, intelligent agents control characters in the virtual environment (virtual humans), playing the roles of locals, friendly and hostile forces, and other mission team members. These agents must support flexible task-oriented collaboration and naturalistic face-to-face communication between trainees and virtual characters, requiring a broad integration of motor skills, problem solving, gestures, facial expressions, and language. Fig. 9 shows the trainees view of the environment, which is projected on an 8×30 in. screen, rendering the characters life-sized. In the current scenario, the student is in charge of a peacekeeping mission when his unit has a collision with a civilian vehicle, seriously injuring a boy. The trainee can speak with his platoon sergeant, a medic and the victim's mother, all controlled by intelligent agents.[11]

The MRE requires plausible emotional responses to meet its training objectives. A key aspect of leadership training is recognizing when one's subordinates are under stress and judging their advice and responses accordingly. Further, non-verbal feedback is a key way that leaders assess the impact of their orders. Indeed, there is anecdotal evidence that the lack of non-verbal feedback in current training simulation technology leads to an unnatural distance between trainees and their simulated subordinates, with the result that they learn to give "inhuman" orders.[12] MRE also requires plausible emotional responses to meet trainee social expectations. In such an emotionally charged scenario, trainees expect the characters to respond with some emotion to the extent that they will misinterpret certain behaviors as emotional, even when they are not. For example, in some of our early evaluations, the system would exhibit long pauses in response to trainee questions due to inefficiencies in the speech processing software. Trainees would interpret these pauses as emotional turmoil in the character, leading them to wonder what information was being withheld from them. One possible resolution to these requirements, creating emotional behaviors off-line, was impractical as the other technology used in MRE generates behavior dynamically

[11] For a video clip illustrating the interactivity that MRE supports, see www.ict.usc.edu/~gratch/media/MRE.mov.

[12] For example, in certain "no-win" exercises held at the School for Command Preparation, trainees would casually order units to sacrifice themselves with complete disregard for how such orders would impact the morale of units in a real situation (personal communication).

at run-time, thus demanding the type of detailed emotional modeling supported by EMA.

An early prototype of MRE was built using the Steve agent architecture (Rickel & Johnson, 1999), which did not model the impact of emotion. We incorporated EMA into Steve, which is also implemented in Soar, to address the application requirements.[13] The resulting agents conform to the basic blackboard architecture laid out in Fig. 2. Each agent incorporates several reasoning functions implemented as cognitive operators that operate over of features of Soar's working memory. These include: a planning, re-planning and execution operators based on Steve (Rickel & Johnson, 1999) that, in MRE, operates over a task model of 40 STRIPS operators and about 50 state predicates; dialogue management operators that interpret and generate speech and dialogue acts, maintains an explicit dialogue state, and support task-related negotiation and multi-party conversation (Traum, Rickel, Gratch, & Marsella, 2003b); natural language generation operators that generate emotionally influenced speech (Fleischman & Hovy, 2002); and perceptual operators that models perceptual attention. In addition, several reasoning modules exist outside the core Soar architecture including speech recognition and understanding modules, an expressive speech synthesizer (Johnson et al., 2002), a gesture generation module, and a procedural animation system that supports facial expressions and communicative gestures. The BEAT system is used to synchronize the gestures with the production of phonemes and visemes (Cassell, Vilhjálmsson, & Bickmore, 2001).

Although our initial focus was on generating appropriate facial expressions and gestures, Steve's blackboard implementation allowed us to tightly integrate EMA into the incremental processing of individual modules. Appraisal essentially provided a uniform currency for comparing alternative actions that could be applied to all system operators. Following the discussion in Section 4.4, the coping mechanism associates emotion-relevant information with cognitive operators, and uses this information to bias processing:

- The understanding module uses the max-interpretation object to resolve ambiguous speech references. For example, if the trainee arrives on scene and asks "what happened", many events could satisfy that utterance. "You arrived, sir", would be the response if using a heuristic favoring the most recent event, whereas "there was an accident" is the response favored by preferring the most emotionally intense potential referent.
- The planning module uses appraisals to guide action and plan selection. If there are multiple courses of action for achieving a goal, the agent will prefer the alternative with the largest positive and least negative affect. Additionally, certain coping strategies result in new intentions, which the planner will attempt to satisfy, for example, by asserting intentions over tasks and sub-goals in service of this new intention.
- The dialogue manager uses the output of coping strategies to decide when to take initiative over the dialogue. For example, if a planful strategy forms an intention to act, and that action requires communication, then the agent will attempt to initiate a dialogue to satisfy this intention.
- Speech generation is informed by any coping strategies associated with the response. For example, if the Sergeant decides to cope with the accident by shifting blame to the civilian driver of the other car, this strategy informs how the accident is described ("she rammed into us" as opposed to "there was an accident").
- Behavior generation is informed by the coping. The max emotion of each coping frame is directly expressed through facial expressions and we associate physical behaviors with the execution of each coping strategy. For example, if the Sergeant copes with the accident by shifting blame, he nonverbally expresses anger toward the driver of the other car using a shake of his head and a dismissive wave of his hand towards the driver.

This is not to say that other mechanisms couldn't lead to similar effects, for example, simple decision theory could inform planning, however using a

[13] See (Marsella et al., 2003) for more on the motivations and issues arising from this integration.

single mechanism can simplify the design of these components.

The MRE system is currently undergoing formal evaluations with military cadets. To date, the system has been evaluated by fifteen West Point and ROTC cadets and, from the perspective of the emotional model, initial results are encouraging, with a majority of the cadets rating the behaviors as quite natural. Two more direct evaluations of the realism and social impact of the emotional model are currently ongoing (Gratch & Marsella, 2004a).

6. General discussion

EMA provides a general and comprehensive model of the processes underlying emotion. Prior approaches have either relied on domain-specific appraisal rules (Elliott, 1992) or modeled fewer appraisal distinctions (El Nasr et al., 2000; Moffat & Frijda, 1995; Neil Reilly, 1996). In particular, we feel it is the first process model that explains how the appraisal of an event can change over time (by tying appraisal to an interpretation that can change with further inference). It is the first comprehensive attempt to model the range of human coping strategies, including strategies that are quite unique for AI research. It is also one of the most comprehensive integrations of an appraisal model with other reasoning capabilities including planning, natural language processing, and nonverbal behavior. Here we discuss a number of limitations and possible extensions to the model and address a number of issues raised by our implementation.

6.1. Limitations and extensions

The existing representation of the causal interpretation is sufficient to model a number of appraisal variables but some will involve further consideration. Unexpectedness could be straightforwardly incorporated as the task network already encodes expectations over what will happen in the future (outcomes that are not predicted by currently executing actions would be deemed unexpected). Reasoning about causal attributions (and social reasoning in general) is currently impoverished. For example, reasoning about power involves representing authority and role relationships across agents. These distinctions are actually being made but the auxiliary reasoning modules underlying our Mission Rehearsal implementation, have not as yet been fully integrated into our appraisal model. See (Mao & Gratch, 2003) for some early work along these lines. Reasoning about urgency requires some representation of time, which is not represented in our model, although it is a concern of many decision-theoretic planning systems and some limited temporal reasoning could be easily incorporated into EMA. Finally, to provide a full treatment of social appraisals the system must extend the second-order representations of desire (Dr. Tom does not desire to provide morphine but he believes Jimmy does) to cover modal belief and intention as well (I believe/intend *P* but she doesn't), along the lines suggested by logics of belief and intention.

A potential limitation of EMA from the psychological perspective is its emphasis on goals and goal processing. Psychological theories have identified a span of motivationally related concepts ranging from basic needs and drives to highly abstract concepts such as ego-involvement (Lazarus, 1991). Although the focus of EMA has been on goals as physical states, this wider range of motivational constructs could be mapped into our basic representational scheme. Indeed, components of ego identity often act as goals or preconditions (albeit at a higher-level of abstraction). So if a person's self-concept involves that they are a good singer, they may take certain actions to establish or maintain this identity (e.g., take singing lessons), and these traits may enable subsequent behavior or other life goals (e.g., becoming a rock-star). In this we adopt the view argued by Ellsworth and Scherer (2003). One caveat is that the connection between events and ego-concepts seems weaker (emphasizing partial-goal attainment) while the connection between events and task goals (at least for the domains explored by planning researchers) are stronger (emphasizing all-or-nothing goal attainment).

Other theories, for example, the theory of Ortony, Clore and Collins, make a strong distinction between goals, standards and preferences that we have not adopted in EMA. The appraisals handled by EMA fall largely under the category of goal processing, while standards are posited to characterize social taboos or norms. For example, fornication may satisfy a personal goal but violate a social standard. Our approach is to represent social standards by assigning positive (negative) utility over states or actions that uphold (violate) the standard. Thus, fornication could be represented as having a desirable effect but also an undesirable side-effect (e.g., alienation from one's peers). We have found this sufficient in practice, but it ducks the question of how one makes the mapping between standards and these goals and states, nor does one have an explicit representation of the standard that might inform dialogue or coping strategies. It also limits the model's ability to make distinctions between certain emotional states that other theories support (e.g., guilt from shame). One approach is to make use of the fact that EMA appraises situations from multiple perspectives. For example, shame might arise if others appraise one's act as blameworthy but guilt only arises only when the individual also shares that assessment.

The emotion focusing mechanism implements a direct connection between appraisals and coping strategies, however, a number of psychologists argue that the connection is far less direct (Clore, Schwartz, & Conway, 1994). For example, anger at a boss may lead to an angry outburst with a spouse over a minor annoyance later in the day, even though there is no causal relation between what caused the anger and the later outburst. In such behavior, there is not a clear causal connection between the appraised event and the response. Rather, the emotion seems to persist and impact later behavior. One possible way to model this behavior is by adjusting the focus mechanism to make use of mood, which encodes a more global and persistent notion of emotional state. We could then adjust the focus mechanism to emphasize certain appraisals that are congruent with the current mood.

6.2. Behavioral consistency

Behavioral consistency is one of the key challenges facing the design of interactive lifelike agents. Such agents involve a variety of reasoning functions (perception, planning, natural language processing). As each of these functions has an associated behavioral manifestation, the problem becomes one of coordinating amongst these functions and conveying an outward appearance of a single coherent individual. Emotion (more specifically, an emotions underlying appraisal) is often posited as playing a key role in addressing this problem in natural organisms, and we claim it can play a similar role in agent design. Our approach is essentially a blackboard model. The causal interpretation summarizes information from various auxiliary reasoning modules. Perception alters the truth-value of directly perceivable states. Planning updates future possible actions and the probabilities associated with state achievement. Dialogue with other agents can update the agent's interpretation of truth-values, intentions, as well as the past or future actions of other agents. Other auxiliary modules like plan recognition could similarly summarize their results through the causal interpretation. By appraising such a unified data structure, appraisal and coping are able to impose a coherent interpretation over these various processes, and use this interpretation to inform behavior and the subsequent direction of the auxiliary reasoning modules.[14]

Beyond enforcing consistency across processing modules, there is the issue of enforcing consistency across time. Here there is a tension between dynamics and inertia. We want to avoid the extremes of an agent whose emotions are always caught up in the moment or an agent who remains unflappable even in the most extreme circumstances. EMA strikes a particular balance between these two extremes – by tying emotional expression to activated subsets of the causal interpretation – but this balancing act is more of an art than a science. If these subsets are defined too narrowly, the agent may seem to jump too rapidly from emotion to emotion. User studies and considerably more

[14] Alternatively, behavioral consistency could be realized by explicitly modeling and measuring the coherency between the network of beliefs, goals and behaviors (Thagard, 2000; Nerb & Spada, 2001).

experience with the system will be needed before we can make any concrete claims about appropriate emotional dynamics.

However, because EMA ties appraisal and coping to cognitive operations on the causal interpretation, the issue of consistency over time may already be addressed, at least partially. First, these cognitive operations don't randomly access different parts of the causal interpretation, but explore the interpretation through a coherent strategy (something that coping strategies should enforce). Indeed, current auxiliary modules (such as planning) tend to walk through the causal interpretation in a systematic manner (e.g., working backward from a top-level goal), and reflecting the dynamics of such mental processes was part of the motivation of activating subsets of appraisals. Second, the causal interpretation incorporates past, present and future. This, combined with the appraisal and coping processes that operate over the causal interpretation, forces consistency over time.

6.3. Relationship to classical decision theory

EMA draws heavily on decision theory but it departs significantly in several respects, suggesting potentially useful extensions to classical decision-theoretic approaches. A purely decision-theoretic approach argues for a view where desirability and likelihood are the only dimensions along which events should be appraised. In contrast, appraisal theories posit that additional dimensions (e.g., attributions of blame or credit) are critical for characterizing human behavior. In this sense, decision theory provides a useful, but incomplete, set of constructs for modeling human behavior, and may be incomplete from the perspective of modeling intelligent behavior in general, independent of its humanness.

EMA also differs in how it combines utility values to inform behavior. Classical decision theory argues that actions should be selected on the basis of their expected utility. For example, Dr. Tom's choice to give morphine should be based on the expected utility of all of the actions outcomes (the utility of each outcome weighted by its probability). Giving morphine has relatively small expected utility as the strong positive and negative consequences of the action cancel each other out. In contrast, EMA appraises each outcome separately, generating strong positive and negative appraisals. Thus, EMA distinguishes actions with strong positive and negative outcomes from actions that have only neutral outcomes, whereas a traditional application of decision theory would treat these actions the same. Further, multiple appraisals are expressed in different proportion depending on the agent's focusing mechanism. For example, if the agent focuses on positive outcomes of giving morphine through its interaction with the user or through its coping strategies, it will tend to display more positive emotions about the procedure. Thus, EMA can mimic standard framing effects whereby how one presents information influences its evaluation (Tversky & Kahneman, 1981), something classical decision-theory cannot account for. Finally, EMA differs significantly from decision theory in that coping strategies act to modify probabilities and utilities which decision theory traditionally treats as constants.

Finally, appraisal theory argues for a re-evaluation of how the concept of rationality is applied to the assessment of human behavior. For example, in using emotion-focused coping strategies, people distort their beliefs for "emotionally convenient" reasons, yet these "irrational" distortions can be highly adaptive, decreasing stress levels, extending life expectancy, enhancing the strength of social relationships. This adaptive nature of emotional behavior may be attributed to the fact that such coping strategies attempt to form a comprehensive response that balances the global physical and social consequences of individual beliefs and decisions. In other words, the common complaint that people are irrational may be more a statement about the artificially narrow inputs to our rational models, rather than the mal-adaptiveness of human decision-making.

6.4. Relationship to models of belief and intention

EMA draws on models of belief, desire and intention to represent an agent's mental state, but it also has implications for that body of work, particularly with respect to the question of when to form or abandon a commitment. Intentions are typically viewed as a commitment to act that constrains subsequent reasoning and behavior. This is illustrated, for example, in the distinction between knowing of a plan verses having a plan, as the latter presumes that subsequent decisions will be consistent with the intention. This notion of commitment is argued to contribute to bounded-decision making, to ease the problem of juggling multiple goals, and coordinate group problem solving.

This raises the question of when to abandon a commitment and EMA suggests a novel solution to the problem. The standard solution is to abandon a commitment if it is inconsistent with an agent's beliefs, but coping strategies like denial complicate the picture, at least with respect to modeling human-like decision making. People can be strongly committed to a belief, even when it contradicts perceptual evidence or their other intentions or social obligations. This suggests that there is no simple criterion for abandoning commitments, but rather one must weight the pros and cons of alternative conflicting commitments. Appraisal and coping provide a possible mechanism for providing this evaluation. Appraisal identifies particularly strong conflicts in the causal interpretation (for example, as when the intention to give morphine violates an intention to preserve life). Coping assesses alternative strategies for resolving the conflict, dropping one conflicting intention or changing some belief so that the conflict is resolved.

6.5. Relationship to natural language processing

Human speech is infused with emotion. On the production side, people encode emotional information in the acoustic properties of their speech, through their choice of words, and even through the structure of their dialogues (for example, anxious or angry individuals may be more inclined to interrupt a speaker or take conversational initiative). On the understanding side, people readily decode such emotional signals, though their own emotional state may significantly impact or bias this decoding process. Our framework suggests how to realize such phenomena throughout the natural language pipeline. Here we briefly consider two aspects that we have explored and implemented.

One challenge in language generation is how to produce language that expresses both the desired information and the desired emotional attitude towards that information. Some work has considered how to generate emotional text given a suitable emotion markup, although these systems did not consider how such a markup might be automatically generated (Bateman & Paris, 1989; Hovy, 1990). In general, these systems choose among a small set of phrases, and within the phrase from a small set of lexical fillers for certain positions of the phrase, where each alternative phrase and lexical item was pre-annotated with an affective value such as *good* or *bad*. The presence of a fine-grained emotion model provides a richer set of distinctions to inform the generation process. Our framework explicitly represents emotional attitudes towards specific states, events, and individuals. Coping strategies, in particular, provide a here-to-fore unexplored means to inform speech generation. For example, if a person involved in an accident wishes to shift blame, they might say "I was rammed" rather than the more neutral "there was an accident". Together, our framework makes available more affective information and facilitates a more nuanced set of expressive alternatives than prior work. This has been demonstrated by the recent work of Fleischman and Hovy that builds on our model (Fleischman & Hovy, 2002).

A challenge in understanding speech is resolving ambiguity and, again, our emotional framework can provide some guidance in addressing this problem. To correctly interpret ambiguous utterances, one must understand what is in linguistic focus. For example, if a policeman arrives at the scene of an accident he might ask "What happened here?" In principle many things have happened. The most common heuristic for modeling linguistic focus is *recency*, which might lead to the factually correct response, "Well, you just drove up". To produce a more appropriate response, a model of focus must account for the fact that people are often focused most strongly on the things that upset them emotionally, such as a traffic accident, which suggests an emotion-based heuristic for determining linguistic focus. When an utterance has multiple possible referents, this heuristic would prefer the referent that generates the greatest emotional charge within the causal interpretation.

6.6. Relationship to intelligent agent design

From a pure software engineering standpoint, EMA has a number of advantages over prior models of emotion or appraisal. The model is completely domain-independent and domain-specific information is encoded using the same knowledge representation schemes used by many autonomous agent systems (e.g., STRIPS operators, utility functions). Although EMA has focused on planning and dialogue as the source of information in the causal interpretation, the knowledge representation is quite general and, as a blackboard architecture, it is relatively straightforward to introduce other processing modules to increase the systems capabilities. EMA's implementation in Soar also facilitates this integration. Soar's working memory serves as the blackboard and reasoning modules (e.g., planning, dialogue management) post their intermediate results to the blackboard. Motivated by psychological theories of cognition, Soar enforces a strict serial bottleneck over operator execution. Different operations must compete for processing resources. Soar's operator preference scheme allows one to easily assert preferences over these competing operators. In contrast to serial cognitive operations, Soar also allows a set of elaboration rules that fire in parallel and are triggered automatically by changes to the blackboard. We make use of this distinction to capture the psychological distinction between deliberate cognitive processes and automatic appraisals. Reasoning modules are implemented via sets of Soar operators whereas appraisal is implemented by elaboration rules. Coping strategies act by posting preferences over different cognitive operators.

We fully expect that the structure of the causal interpretation will have to be generalized to accommodate some of these additions. For example, a number of multi-agent systems have stronger architectural support for reasoning about distributed problem solving. For example GPGP (Decker & Lesser, 1992) maintains different data structures for individual vs. joint plans. Dialogue systems maintain distinctions between grounded (mutually agreed upon) and ungrounded plans and assertions (Traum & Rickel, 2002). The psychological literature also suggests that people represent detailed information about the motivation lying behind other individual's plans (e.g., were they coerced, did they foresee the outcome, did they have an alternative), and use this motivation in assigning blame and credit (Shaver, 1985; Weiner, 1986). While the current structure of the causal interpretation supports some of these distinctions, others will require modification of the data structures (for example the current system has an impoverished model of belief).

6.7. Physiological substrate

At the beginning of this article, we quoted John Searle, "You don't get emotions by manipulating 1s and 0s". Our work addresses this challenge but only in part. We have modeled the cognitive components of emotion. This ignores, for example, the physiological and biochemical components of emotion that to date has not been a focus of our work. In particular, our model forms emotions in a fashion that is only loosely coupled to the physical processes that realize the computation. In human emotions, one can argue, as Searle does, that there is a strong coupling between cognitive processes and underlying physiological processes.

In contrast to the John Searle quote, we also quoted Marvin Minsky, "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions." The model we have presented has shown how emotions can be computationally modeled. Further, consistent with Minsky's view, the model emphasizes a central role for emotions in intelligent behavior. Key issues for our work moving forward will be whether the physiological components of emotion need to be modeled, which components need to be modeled and how they can be modeled.

Assuming an answer to these issues, another challenge here would be how to model the relation of the physiological components on EMA's appraisal and coping processes. One speculative answer would be to mediate the relation through mood. Physiology would impact mood which in turn would bias appraisal and coping, in particular influencing assessments of coping potential. Additionally, appraisal could in turn modify mood, much as it currently does in EMA.

6.8. Evaluation

Given the pervasive influence emotions have over behavior, evaluating the effectiveness of such a general architecture presents some unique challenges. Emotional influences are manifested across a variety of levels and modalities. Emotion is often attributed to others in response to telltale physical signals: facial expressions, body language, and certain acoustic features of speech. But emotion is also conveyed through patterns of thought and coping behaviors such as wishful thinking, resignation, or blame-shifting. Unlike many phenomena studied by cognitive science, emotional responses are also highly variable, differing widely both within and across individuals depending on non-observable factors like goals, beliefs, cultural norms, etc. And unlike work in decision making, there is no accepted normative model of emotional responses or their dynamics that we can use as a gold standard for evaluating techniques.

Researchers in the lifelike-character community evaluation have relied largely on the concept of "believability" in demonstrating the effectiveness of a technique. A human subject is allowed to interact with a system or see the result of some system trace, and is asked how believable the behaviors appear; it is typically left to the subject to interpret what is meant by the term. One obvious limitation with this approach is that there seems to be no generally agreed definition of what "believability" means, how it relates to other similar concepts such as realism, or how to operationalize its evaluation. For example, in a health-intervention application developed by one of the authors, stylized cartoon animation was judged to be highly believable even though it was explicitly designed to be unrealistic along several dimensions (Marsella et al., 2003).

In evaluating our work, we are attempting to move beyond the concept of believability and instead ask more focused questions that we can more carefully evaluate. Here we briefly preview two studies currently underway at our lab that illustrate this approach. In the first study, we address the question of behavior generation: does the model generate behavior that is consistent with how people actually behave, specifically with regard to how emotion and coping unfold over time. In the second, we address the question of behavioral influence: do the generated behaviors have the same social influence on a human subject that one person's emotion has on another person. In other words, (1) does our computational model create the right cognitive dynamics and (2) does it have the right social impact.

For evaluating cognitive dynamics, we are attempting to fit our model to a standard instrument used in the clinical psychological evaluation of a person's emotional and coping response to stressful situations, and in particular, how these responses evolve over time. In the Stress and Coping Process Questionnaire (Perrez & Reicherts, 1992), a subject is presented a stereotypical situation, such as an argument with their boss. They are asked how they would respond emotionally and how they would cope. They are then given subsequent updates on the situation and asked how their emotions/coping would dynamically unfold in light of systematic variations in both expectations and perceived sense of control. Based on their evolving pattern of responses, subjects are scored as to how closely their reactions correspond to those of normal healthy adults. In our evaluation, we encode these evolving situations in EMA's domain language, run the scenarios, and compare EMA's appraisals and coping strategies to the responses indicated by the scale. Using such a scale has several advantages. First, the situations in the instrument were formalized by someone outside our research group, and thus constitute a fairer test of the approach's generality than what is often performed (though we are clearly subject to bias in our selection of a particular instrument). Second, by formalizing an evolving situation, this instrument directly assesses the question of emotional dynamics, rather than single situation-response pairs typically considered in evaluations. Finally, the exercise of encoding situations into a domain theory acceptable by our model clearly delineates the limits of the model, for example, in terms of what aspects of the situation were naturally expressed and which could only be handled outside the model.

For evaluating the social impact of our model, we are initially focusing on the phenomena of social referencing. whereby people, when presented with an ambiguous decision, are influenced by appraisals of others (Campos, 1983). In our evaluation, we assess the ability of our model to induce social referencing in human subjects in the context of the Mission Rehearsal Exercise. Subjects, while interacting in a virtual environment with virtual characters, are forced to make a decision between two courses of action where the correct decision is ambiguous. Across two experimental conditions, we vary which decision the virtual characters prefer and this preference is expressed through non-verbal responses of the characters. The hypothesis is that human subjects will be influenced by this non-verbal behavior in a way that is consistent with the phenomenon of social referencing.

7. Conclusion

In this article, we have outlined a general computational framework of appraisal and coping as a central organizing principle for human-like autonomous agents. Three aspects of human emotional behavior have heavily influenced our approach to computational modeling of emotion. First, emotion is an organizing principle of human behavior, both influenced by, and influencing in return, a wide range of cognitive and physical behaviors. Second, emotion is central to adaptive behavior and adaptive behavior is more than just immediate actions in the world. People adapt their beliefs, goals and plans and emotion plays a central role here as well. Finally, emotion, and the behaviors it influences, operates over the past, present and future. People have beliefs about past events, emotions about those events and can alter those emotions by altering the beliefs. Similarly, people have beliefs about the future, experience emotions through those beliefs, and can seek to change those beliefs in various ways. Emotion, because it operates over both the past and future, provides a temporal consistency to behavior and beliefs.

These requirements lead us to a computational model of emotion that is tied to an individual's causal interpretation of the world. This interpretation provides a rich context that ties together emotional processes with other internal and external behaviors and allows them to interact within a uniform representation of the past, present and future. As a result the appraisal model described here is unique among computational models of emotion both in its ability to be influenced by, and in turn to influence, other cognitive processes. As significant, the emotion model is unique in its ability to derive emotions from beliefs about the past and in coping's ability to potentially influence those beliefs. Furthermore, by capturing these capabilities through domain-independent reasoning, and by building on standard AI processes and representations, the approach can be readily incorporated into a number of autonomous agent systems, and is of potential use, not only for inferring emotional state, but also for informing a number of the behaviors that must be modeled by virtual humans such as facial expressions, dialogue management, planning, reacting, and social understanding. Thus, we expect appraisal models such as this to inform the ever-increasing range of applications where human and computers must grapple with the daunting task of mutual coexistence and understanding.

Acknowledgements

This work was funded by the Department of the Army under contract DAAD 19-99-D-0046. We thank our many colleagues who are contributing to the MRE project especially Jeff Rickel, David Traum, Ed Hovy, Michael Fleishman and Lewis Johnson. John Laird, Aaron Sloman, Paul Rosenbloom and Wenji Mao provided valuable feedback on an earlier draft. The anonymous reviewers also provided many helpful comments.

Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

Agre, P., & Chapman, D. (1987). *Pengi: An implementation of a theory of activity.* Paper presented at the national conference on artificial intelligence.

Ambros-Ingerson, J., & Steel, S. (1988). *Integrating planning, execution and monitoring.* Paper presented at the seventh national conference on artificial intelligence, St. Paul, MN.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

André, E., Rist, T., Mulken, S. v., & Klesen, M. (2000). The automated design of believable dialogues for animated presentation teams. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 220–255). Cambridge, MA: MIT Press.

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

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Bateman, J. A., & Paris, C. L. (1989). *Phrasing a text in terms the user can understand.* Paper presented at the 11th international joint conference on artificial intelligence, Detroit, MI.

Bates, J., Loyall, B., & Reilly, W. S. N. (1991). Broad agents. *Sigart Bulletin, 2*(4), 38–40.

Blythe, J. (1999). Decision theoretic planning. *AI Magazine, 20*(2), 37–54.

Boutilier, C., Dean, T., & Hanks, S. (1999). Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research, 11*, 1–94.

Bratman, M. (1990). What is intention? In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: MIT Press.

Campos, J. J. (1983). The importance of affective communication in social referencing: A commentary on Feinman. *Merrill-Palmer Quarterly, 29*, 83–87.

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality Psychology, 56*(2), 267–283.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000). Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 29–63). Boston: MIT Press.

Cassell, J., Vilhjálmsson, H., & Bickmore, T. (2001). *BEAT: The behavior expressive animation toolkit*. Paper presented at the SIGGRAPH, Los Angeles, CA.

Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence, 32*, 333–377.

Clore, G., Schwartz, N., & Conway, M. (1994). Affect as information. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 121–144).

Clore, G. L., & Gasper, K. (2000). Feeling is believing: Some affective influences on belief. In N. Frijda, A. S. R. Manstead, & S. Bem (Eds.), *Emotions and beliefs: How feelings influence thoughts* (pp. 10–44). Paris: Cambridge University Press.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotion* (2nd ed., pp. 91–115). New York: Guilford Press.

Costa, P., Somerfield, M., & McCrae, R. (1996). Personality and coping: A reconceptualization. In M. Zeidner & N. Endler (Eds.), *Handbook of coping*. New York: Wiley.

Damasio, A. R. (1994). *Descartes' error: Emotion reason and the human brain*. New York: Avon Books.

de Kleer, J., & Brown, J.S., (1982). *Foundations of envisioning.* Paper presented at the national conference on artificial intelligence, Pittsburgh, PA.

Decker, K., & Lesser, V. (1992). Generalizing the partial global planning algorithm. *International Journal of Intelligent and Cooperative Information Systems, 1*(2), 319–346.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska symposium on motivation* (pp. 207–283). Lincoln, Nebraska: University of Nebraska Press.

El Nasr, M. S., Yen, J., & Ioerger, T. (2000). FLAME: Fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems, 3*(3), 219–257.

Elliott, C., (1992). The affective reasoner: A process model of emotions in a multi-agent system (Ph.D. Dissertation No. 32). Northwestern, IL: Northwestern University Institute for the Learning Sciences.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of the affective sciences* (pp. 572–595). New York: Oxford University Press.

Fikes, R., & Nilsson, N. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence, 2*(3–4), 189–208.

Fleischman, M., & Hovy, E. (2002). *Emotional variation in speech-based natural language generation.* Paper presented at the international natural language generation conference, Arden House, NY.

Frank, R. (1988). *Passions with reason: The strategic role of the emotions.* New York: W.W. Norton.

Frijda, N. (1987). Emotion cognitive structure and action tendency. *Cognition and Emotion, 1*, 115–143.

Frijda, N. H., & Zeelenberg, M. (2001). Appraisal: What is the dependent? In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 141–156). Oxford: Oxford University Press.

Gratch, J. (2000). *Émile: Marshalling passions in training and education.* Paper presented at the fourth international conference on intelligent agents, Barcelona, Spain.

Gratch, J. (2002). Socially situated planning. In K. Dauterhahn, A. H. Bond, L. Cañamero, & B. Edmonds (Eds.), *Socially intelligent agents – Creating relationships with computers and robots* (pp. 181–188). Norwell, MA: Kluwer Academic Publishers.

Gratch, J., & Marsella, S., 2001. *Tears and fears: Modeling emotions and emotional behaviors in synthetic agents.* Paper presented at the fifth international conference on autonomous agents, Montreal, Canada.

Gratch, J., & Marsella, S. (2003). Fight the way you train: The role and limits of emotions in training for combat. *Brown Journal of World Affairs, X*(1).

Gratch, J., & Marsella, S. (2004a). *Evaluating a general model of emotional appraisal and coping.* Paper presented at the AAAI symposium on architectures for modeling emotion: Cross-disciplinary foundations, Palo Alto, CA.

Gratch, J., & Marsella, S. (2004b). *Technical details of a domain independent framework for modeling emotion*. Technical Report ICT-TR-04-2004. From www.ict.usc.edu/~gratch/EMA_Details.pdf.

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems* (July/August), 54–61.

Grosz, B., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence, 86*, 2.

Hill, R., Chen, J., Gratch, J., Rosenbloom, P., & Tambe, M. (1997). *Intelligent agents for the synthetic battlefield.* Paper presented at the joint proceedings of the fourteenth national conference on artificial intelligence and the ninth conference on innovative applications of artificial intelligence (AAAI/IAAI97), Providence, RI.

Hovy, E. H. (1990). Pragmatics and natural language generation. *Artificial Intelligence, 43*(2), 153–198.

Johnson, W. L., Narayanan, S., Whitney, R., Das, R., Bulut, M., & LaBore, C. (2002). *Limited domain synthesis of expressive military speech for animated characters.* Paper presented at the 7th international conference on spoken language processing, Denver, CO.

Kirby, L., & Smith, C. (1996). *Freaking, quitting, and staying engaged: Patterns of psychphysiological response to stress.* Paper presented at the ninth conference of the international society for research on emotions, Toronto, Ontario.

Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R. (2001). Relational meaning and discrete emotions. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 37–67). Oxford: Oxford University Press.

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

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion, 14*, 473–493.

Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Instruction, 9*(1–2), 1–44.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic and emotive communication in animated pedagogical agents. In J. Cassell, S. Prevost, J. Sullivan, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 123–154). Cambridge, MA: MIT Press.

Lisetti, C., & Gmytrasiewicz, P. (2002). Can a rational agent afford to be affectless? A formal approach. *Applied Artificial Intelligence, 16*, 577–609.

Mao, W., & Gratch, J. (2003). *The social credit assignment problem.* Paper presented at the international working conference in interactive virtual agents, Kloster Irsee, Germany.

Marsella, S., & Gratch, J. (2001). *Modeling the interplay of plans and emotions in multi-agent simulations.* Paper presented at the cognitive science society, Edinburgh, Scotland.

Marsella, S., & Gratch, J. (2002). *A step toward irrationality: using emotion to change belief.* Paper presented at the First International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, Italy.

Marsella, S., & Gratch, J. (2003). *Modeling coping behaviors in virtual humans: Don't worry, be happy.* Paper presented at the Second International Joint Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia.

Marsella, S., Gratch, J., & Rickel, J. (2003). Expressive behaviors for virtual worlds. In H. Prendinger & M. Ishizuka (Eds.), *Life-like characters tools, affective functions and applications*. Berlin: Springer.

Marsella, S., Johnson, W. L., & LaBore, C. (2000). *Interactive pedagogical drama.* Paper presented at the fourth international conference on autonomous agents, Montreal, Canada.

Marsella, S., Johnson, W. L., & LaBore, C. (2003). *Interactive pedagogical drama for health interventions.* Paper presented at the artificial intelligence in education, Sydney, Australia.

Mele, A. R. (2001). *Self-deception unmasked.* Princeton, NJ: Princeton University Press.

Minsky, M. (1986). *The society of mind.* New York: Simon and Schuster.

Moffat, D., & Frijda, N. (1995). *Where there's a Will there's an agent.* Paper presented at the workshop on agent theories, architectures and languages.

Neal Reilly, W. S. (1996). Believable social and emotional agents (Ph.D Thesis No. CMU-CS-96-138). Pittsburgh, PA: Carnegie Mellon University.

Nerb, J., & Spada, H. (2001). Evaluation of environmental problems: A coherence model of cognition and emotion. *Cognition and Emotion, 15*, 521–551.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Oatley, K., & Johnson-Laird, P. N. (1987). Cognitive theory of emotions. *Cognition and Emotion, 1*(1).

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

Peacock, E., & Wong, P. (1990). The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal. *Stress Medicine, 6*, 227–236.

Pearl, J. (2002). Reasoning with cause and effect. *AI Magazine, 23*(1), 95–112.

Penley, J., & Tomaka, J. (2002). Associations among the Big Five emotional responses and coping with acute stress. *Personality and Individual Differences, 32*, 1215–1228.

Perrez, M., & Reicherts, M. (1992). *Stress coping and health*. Seattle WA: Hogrefe and Huber Publishers.

Poggi, I., & Pelachaud, C. (2000). Emotional meaning and expression in performative faces. In A. Paiva (Ed.), *Affective Interactions: Towards a New Generation of Computer Interfaces (pp. 182–195)*. Berlin: Springer-Verlag.

Pollack, M. (1990). Plans as complex mental attitudes. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: MIT Press.

Reiter, R. (1987). A logic for default reasoning. In M. Ginsberg (Ed.), *Readings in nonmonotonic reasoning*. Los Altos, CA: Morgan Kaufmann.

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception cognition and motor control. *Applied Artificial Intelligence, 13*, 343–382.

Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., & Swartout, W. (2002). Toward a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems* (July/August), 32–38.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67*(2), 206–221.

Rothbaum, B. O., Hodges, L. F., Alarcon, R., Ready, D., Shahar, F., Graap, K., et al. (1999). Virtual environment exposure therapy for PTSD Vietnam veterans: A case study. *Journal of Traumatic Stress*, 263–272.

Russell, S., & Wefald, E. (1989). *On optimal game-tree search using rational meta-reasoning.* Paper presented at the eleventh international joint conference on artificial intelligence, Detroit, MI.

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning, 19*(2), 195–208.

Scherer, K. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317).

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion*. Oxford: Oxford University Press.

Searle, J. (2002, May 7). Quoted in a human touch for machines. *Los Angeles Times*, p. 1.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness.* New York: Springer.

Shaw, E., Johnson, W. L., & Ganeshan, R. (1999). *Pedagogical agents on the web.* Paper presented at the proceedings of the third international conference on autonomous agents, Seattle, WA.

Silverman, B. G. (2002). Human behavior models for game-theoretic agents: Case of crowd tipping. *Cogn. Sci. Q., Fall.*

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*, 29–39.

Sloman, A., & Croucher, M. (1981). *Why robots will have emotions.* Paper presented at the international joint conference on artificial intelligence, Vancouver, Canada.

Smith, C., & Kirby, L. (2000). Consequences require antecedents: Toward a process model of emotion elicitation. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 83–106). Cambridge, MA: Cambridge University Press.

Smith, C., & Kirby, L. (2001). Towards delivering on the promise of appraisal theory. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 121–138). Oxford: Oxford University Press.

Smith, C., & Lazarus, R. (1990). Emotion and adaptation. In Pervin (Ed.), *Handbook of personality: Theory & research* (pp. 609–637). New York: Guilford Press.

Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Paris: Cambridge University Press.

Thagard, P. (2000). *Coherence in thought and action.* Cambridge, MA: MIT Press.

Thomas, F., & Johnston, O. (1995). *The illusion of life: Disney animation.* New York: Hyperion.

Traum, D., & Rickel, J. (2002). *Embodied agents for multi-party dialogue in immersive virtual worlds.* Paper presented at the first international conference on autonomous agents and multi-agent systems, Bologna, Italy.

Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003a). *Negotiation over tasks in hybrid human-agent teams for simulation-based training.* Paper presented at the second international conference on autonomous agents and multi-agent systems, Melbourne, Australia.

Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003b). *Negotiation over tasks in hybrid human-agent teams for simulation-based training.* Paper presented at the international conference on autonomous agents and multiagent systems, Melbourne, Australia.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer.

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