

The Architectural Role of Emotion in Cognitive Systems

Jonathan Gratch and Stacy Marsella
University of Southern California

Abstract

In this chapter, we will revive an old argument that theories of human emotion can give insight into the design and control of complex cognitive systems. In particular, we claim that appraisal theories of emotion provide essential insight into the influences of emotion over cognition and can help translate such findings into concrete guidance for the design of cognitive systems. Appraisal theory claims that emotion plays a central and functional role in sensing external events, characterizing them as opportunity or threats and recruiting the cognitive, physical and social resources needed to adaptively respond. Further, because it argues for a close association between emotion and cognition, the theoretical claims of appraisal theory can be recast as a requirement specification for how to build a cognitive system. This specification asserts a set of judgments that must be supported in order to correctly interpret and respond to stimuli and provides a unifying framework for integrating these judgments into a coherent physical or social response. This chapter elaborates argument in some detail based on our joint experience in building complex cognitive systems and computational models of emotion.

Introduction

To survive in a dynamic, semi-predictable and social world, organisms must be able to sense external events, characterize how they relate to their internal needs (e.g., is this an opportunity or a threat?), consider potential responses (e.g., fight, flight or plan) and recruit the cognitive, physical and social resources needed to adaptively respond. In primitive organisms, this typically involves hard-wired or learned stimulus-response patterns. For sophisticated organisms such as humans, this basic cycle is quite complex and can occur at multiple levels and time scales, involve deliberation and negotiation with other social actors and can utilize a host of mental functions including perception, action, belief formation, planning and linguistic processing. Progress in modeling such complex phenomena depends on a theory of cognitive system design that clearly delineates core cognitive functions, how they interoperate, and how they can be controlled and directed to achieve adaptive ends.

In this chapter, we will revive an old argument that theories of human emotion can give insight into the design and control of complex cognitive systems, and argue that one theory of emotion in particular, appraisal theory, helps identify core cognitive functions and how they can be controlled ([see also Hudlicka, Chapter X](#)). Debates about the benefit of emotion span recorded history and were prominent, as well, in the early days of cognitive science. Early cognitive scientists argued that emotional influences that seem irrational on the surface have important social and cognitive functions that would be required by *any* intelligent system. For example, Simon (Simon, 1967) argued that emotions serve the crucial function of interrupting normal cognition when unattended goals require servicing. Other authors have emphasized how 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 to cooperate in social games such as prison’s dilemma (Frank, 1988). Similarly, “emotional biases” such as wishful thinking may reflect a rational mechanism that is more accurately accounting for certain social costs, such as the cost of betrayal when a parent defends a child despite strong evidence of their guilt in a crime (Mele, 2001). Ironically, after arguing for the centrality of emotion in cognition, Simon and others in the cognitive modeling community went on to develop narrow focused models of individual cognitive functions that assumed away many of the central control problems that emotion is purported to solve.

After some neglect, the question of emotion has again come to the forefront as models have begun to catch up to theory. This has been spurred, in part, by an explosion of interest in integrated computational models that incorporate a variety of cognitive functions (Anderson, 1993; Bates, Loyall, & Reilly, 1991; Rickel et al., 2002). Indeed, until the rise of broad integrative models of mental function, the problems emotion was purported to solve, for example, juggling multiple goals, were largely hypothetical. More recent cognitive systems embody a variety of mental functions and face very real choices on how to allocate resources. A reoccurring theme in emotion research is the role of emotion in addressing such control choices by directing cognitive resources towards problems of adaptive significance for the organism. Indeed, Simon appealed to emotion to explain how his sequential models could handle the multiplicity of motives that underlie most human activity:

The theory explains how a basically serial information processor endowed with multiple needs behaves adaptively and survives in an environment that presents unpredictable threats and opportunities. The explanation is built on two central mechanisms: 1. A goal-terminating mechanism [goal executor]...2. An interruption mechanism, that is, emotion, allows the processor to respond to urgent needs in real time. (Simon, 1967, p. 39)

Interrupts are part of the story, but contemporary emotion research suggests emotion exact a far more pervasive control over cognitive processes. Emotional state can influence what information is available in working memory (Bower, 1991), the subjective utility of alternative choices (see Busemeyer, this volume), and even the style of processing (Bless, Schwarz, & Kimmelmeier, 1996; Schwarz, Bless, & Bohner, 1991). For example, people who are angry or happy tend to perform more shallow inference and are more influenced by stereotypical beliefs where as sad individuals tend to process more deeply and be more sensitive to the true state of the world. These psychological findings are bolstered by evidence from neuroscience underscoring the close connection between emotion and centers of the brain associated with higher-level cognition. For example, studies performed by Damasio and colleagues suggest that damage to ventromedial prefrontal cortex prevents emotional signals from guiding decision making in an advantageous direction, particularly for social decisions (Bechara, Damasio, AntonioDamasio, & Lee, 1999). Other studies have illustrated a close connection between emotion and cognition via the anterior cingulate cortex, a center of the brain often implicated in cognitive control (Allmana, Hakeema, Erwinb, Nimchinskyc, & Hofd, 2001). Collectively, these findings demonstrate that emotion and cognition are closely coupled and suggest emotion has a strong, pervasive and controlling influence over cognition.

We argue appraisal theory (Arnold, 1960; Frijda, 1987; Lazarus, 1991; Ortony, Clore, & Collins, 1988; Scherer, 1984), the most influential contemporary theory of human emotion, can help make sense of the various influences of emotion over cognition and, further, help translate such findings into concrete guidance for the design of cognitive systems. Appraisal theory asserts that emotion plays a central and functional role in sensing external events, characterizing them as opportunity or threats and recruiting the cognitive, physical and social resources needed to adaptively respond. Further, because it argues for a close association between emotion and cognition, the theoretical claims of appraisal theory can be recast as a requirement specification for how to build a cognitive system – it claims a particular set of judgments must be supported in order to correctly interpret and respond to stimuli and provides a unifying framework for integrating these judgments into a coherent physical or social response. This chapter elaborates argument in some detail based on our joint experience in building complex cognitive systems and computational models of emotion.

Computational Appraisal Theory

Appraisal theory is the predominant psychological theory of human emotion and here we argue that it is also the most fruitful theory of emotion for those interested in the design of cognitive systems (Arnold, 1960; Frijda, 1987; Lazarus, 1991; Ortony et al., 1988; Scherer, 1984).¹ The theory emphasizes the connection between emotion and cognition, arguing that emotions are an aspect of the mechanisms by which organisms detect, classify, and adaptively respond to significant changes to their environment. A central tenant is that emotions are associated with patterns of individual judgment that characterize the personal significance of external events (e.g., was this event expected in terms of my prior beliefs? is this event congruent with my goals; do I have the power to alter the consequences of this event?). These judgments involve cognitive processes, including slow deliberative, as well as fast automatic or associative processes.

There are several advantages to adopting an appraisal-theoretic perspective when approaching the problem of cognitive system design. Unlike neuroscience models, appraisal theory is often cast at a conceptual level that meshes well with the level of analysis used in most cognitive systems, as emotions are described in terms of their relationship to goals, plans and problem-solving. In this sense, appraisal theories contrast sharply with categorical theories (Ekman, 1992) that postulate a small set of innate hardwired neuromotor programs that are separate from cognition, or dimensional theories that argue emotions are classified along certain dimensions and make no commitment to underlying mechanism (Russell & Lemay, 2000). Finally, as a paradigm that has seen consistent empirical support and elaboration over the last fifty years, appraisal theory has been applied to a wide range of cognitive and social phenomena, and thus provides the most comprehensive single framework for conceptualizing the role of emotion in the control of cognition.

Appraisal and Coping

Appraisal theory argues that emotion arises from the dynamic interaction of two basic processes: appraisal and coping (Smith & Lazarus, 1990). Appraisal is the process by which a person assesses their overall relationship with its environment, including not only their current condition but past events that led to this state as well as future prospects. Appraisal theory argues that appraisal, although not a deliberative process in of itself, is informed by cognitive processes and, in particular, those process involved in understanding and interacting with the physical and social

environment (e.g., planning, explanation, perception, memory, linguistic processes). Appraisal maps characteristics of these disparate mental processes into a common set of terms called *appraisal variable* (e.g., is this event *desirable*, who *caused* it, what *power* do I have over its unfolding). These variables serve as an intermediate description of the person-environment relationship – a common language of sorts – and are claimed to mediate between stimuli and response (e.g. different responses are organized around how a situation is appraised). Appraisal variables characterize the significance of events from the individual’s perspective. Events do not have significance in of themselves, but only by virtue of their interpretation in the context of an individual’s beliefs, desires and intention, and past events.

Coping refers to how one responds to the appraised significance of events. People are motivated to respond to events differently depending on how they are appraised (Peacock & Wong, 1990). For example, 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 denial or resignation. Appraisal theories often characterize the wide range of human coping responses into two broad classes: *problem-focused coping strategies* attempt to change the environment; *emotion-focused coping strategies* (Lazarus, 1991) involves inner-directed strategies for dealing with emotions, for example, by discounting a potential threat or abandoning a cherished goal. 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 (re-) appraisals. Thus, coping, cognition and appraisal are tightly coupled, interacting and unfolding over time (Lazarus, 1991): an agent experience fear upon perceiving a potential threat (appraisal), which motivates problem solving (coping), which leads to relief upon deducing an effective countermeasure (re-appraisal). A key challenge for any model of this process is to capture these dynamics.

EMA: A Computational Perspective

EMA is a computational model that attempts to concretize the mapping between appraisal theory and cognitive system research (Gratch & Marsella, 2001, 2004, 2005; Marsella & Gratch, 2003).² Given appraisal theory’s emphasis on a person’s evolving *interpretation* of their relationship with the environment, EMA’s development has centered on elucidating the mechanisms that inform this interpretation and how emotion informs and controls the subsequent functioning of these mechanisms. At any point in time, the agent’s current view of the agent-environment relationship is represented in “working memory”, which changes with further observation or inference. EMA treats appraisal as a set of feature detectors that map features of this representation into appraisal variables. For example, an effect that threatens a desired goal is assessed as a potential undesirable event. Coping is cast as a set of control signals that direct the processing of auxiliary reasoning modules (i.e., planning, belief updates, etc.) to overturn or maintain those features that yielded the appraisals. For example, coping could resign the agent to the threat by abandoning the desired goal, or alternatively, it could signal the planning system to explore contingencies. Figure 1 illustrates this perspective on appraisal theory as a mechanism for the control of cognition.

In support of a mechanistic account, we have adopted a strategy of using conventional artificial intelligence reasoning techniques as proxies for the cognitive mechanisms that are claimed to underlie appraisal and coping. Appraisal theory posits that events are interpreted in terms of several appraisal variables, discussed below, that collectively can be seen as a requirement specifi-

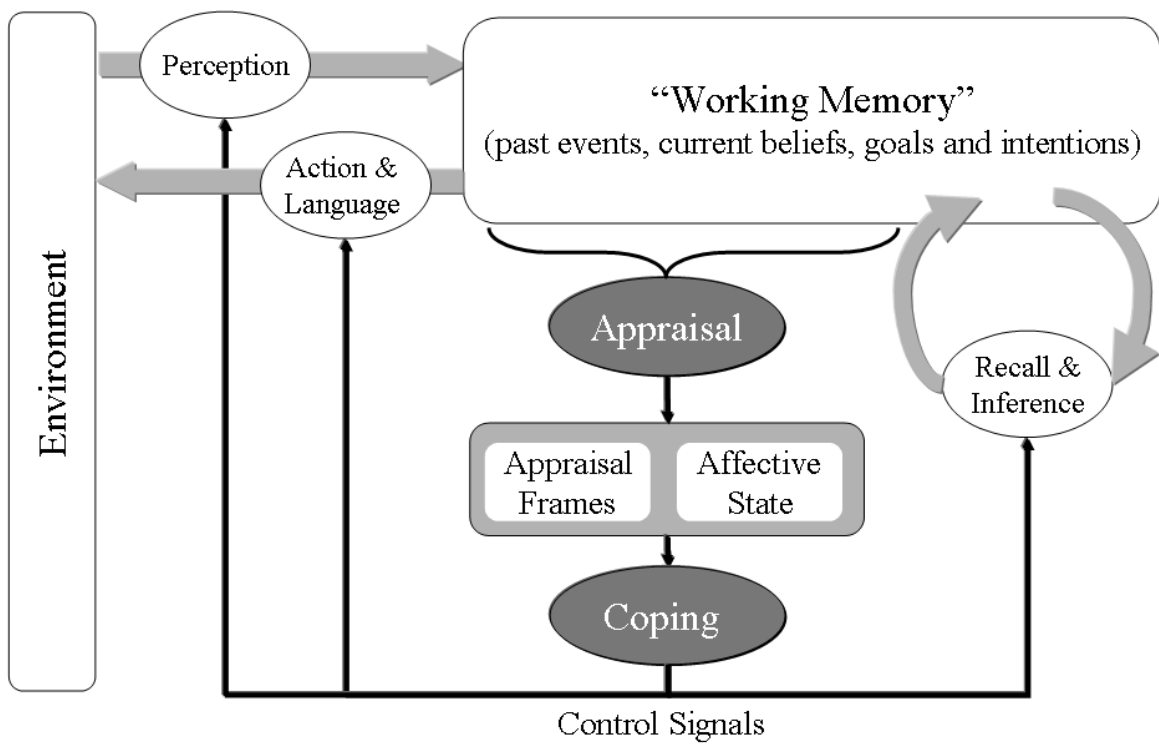


Figure1: A view of emotion as ‘affective control’ over cognitive functions.

cation for the classes of inference a cognitive system must support. This specification is far broader than what is typically supported by conventional artificial intelligence techniques, so to capture this interpretative process within a computational system, we have found it most natural to integrate a variety of reasoning methods. Specifically, we build on the causal representations developed for decision-theoretic planning (Blythe, 1999) and augment them with methods that explicitly model commitments to beliefs and intentions (Grosz & Kraus, 1996; Pollack, 1990). 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 many autonomous agent reasoning techniques (e.g., planning, explanation, natural language processing). The decision-theoretic concepts of utility and probability are crucial for modeling appraisal variables related to the desirability and likelihood of events. Explicit representations of intentions and beliefs are critical for assessing the extent to which an individual deserves blame or credit for their actions, as such attributions involve judgments of intent, foreknowledge and freedom of choice (Shaver, 1985; Weiner, 1995). As we will see, commitments to beliefs and intentions also play a role in modeling coping strategies.

In EMA, the agent’s interpretation of its “agent-environment relationship” is reified in an explicit representation of beliefs, desires, intentions, plans and probabilities (see Figure 2). Following a blackboard-style model, this representation (corresponding to the agent’s working memory) encodes the input, intermediate results and output of reasoning processes that mediate between the agent’s goals and its physical and social environment (e.g., perception, planning, explanation, and natural language processing). We use the term *causal interpretation* to refer to this collection of data structures to emphasize the importance of causal reasoning as well as the interpretative (subjective) character of the appraisal process. Figure 2 illustrates an instance of this data structure in which an agent has a single goal (affiliation) that is threatened by the recent departure of a friend (the past “friend departs” action has one effect that deletes the “affiliation” state).

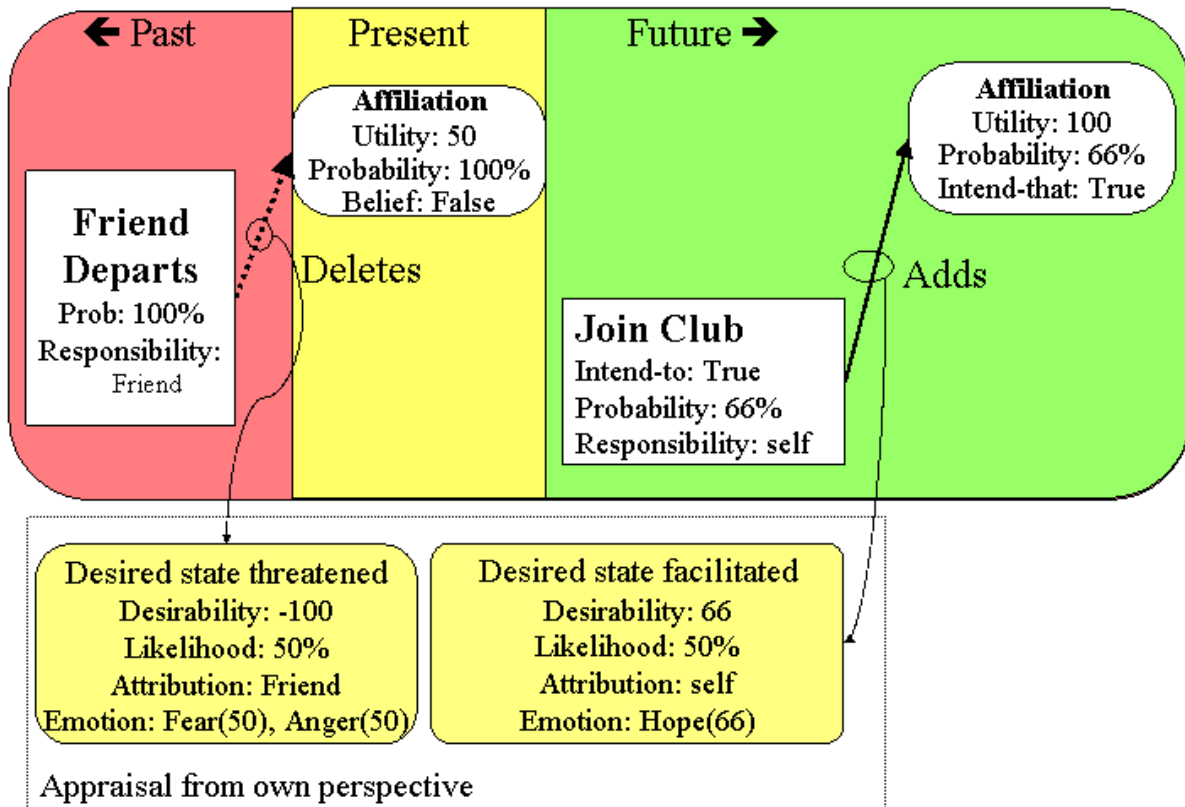


Figure2: An instance of a causal interpretation and associated appraisal frames.

This goal might be re-achieved if the agent joins a club. Appraisal assesses each case where an act facilitates or inhibits some proposition in the causal interpretation. In the figure, the interpretation encodes two “events,” the threat to the currently satisfied goal of affiliation, and the potential re-establishment of affiliation in the future. Associated with each event in the causal interpretation is an *appraisal frame* that summarizes, in terms of appraisal variables, its significance to the agent.

Each event is characterized in terms of appraisal variables by domain-independent functions that examine the syntactic structure of the causal interpretation:

- Perspective: from whose viewpoint is the event judged?
- Desirability: what is the utility of the event if it comes to pass, from the perspective taken (e.g., does it causally advance or inhibit a state of some utility)? The utility of a state may be intrinsic (agent X attributes utility Y to state Z) or derived (state Z is a precondition of a plan that, with some likelihood, will achieve an end with intrinsic utility).
- Likelihood: how probable is the outcome of the event? This is derived from the decision-theoretic plan.
- Causal attribution: who deserves credit or blame? This depends on what agent was responsible for executing the action, but also involves epistemic considerations such as intention, foreknowledge and coercion (see (Mao & Gratch, 2004)).
- Temporal status: is this past, present, or future?

- **Controllability:** can the outcome be altered by actions under control of the agent whose perspective is taken? This is derived by looking for actions in the causal interpretation that could establish or block some effect, and that are under control of the agent whose perspective is being judged (i.e., agent X could execute the action).
- **Changeability:** can the outcome be altered by external processes or some other causal agent? This involves consideration of actions believed available to others as well as their intentions.

Each appraised event is mapped into a discrete emotion instance of some type and intensity, following the scheme proposed by Ortony et al (Ortony et al., 1988). A simple activation-based focus of attention model computes a current emotional state based on most-recently accessed emotion instances.

Coping determines how one responds to the appraised significance of events. Coping strategies are proposed to maintain desirable or overturn undesirable in-focus emotion instances. Coping strategies essentially work in the reverse direction of appraisal, identifying the precursors of emotion in the causal interpretation that should be maintained or altered (e.g., beliefs, desires, intentions, and expectations).

Strategies include:

- **Action:** select an action for execution
- **Planning:** form an intention to perform some act (the planner uses intentions to drive its plan generation)
- **Seek instrumental support:** ask someone that is in control of an outcome for help
- **Procrastination:** wait for an external event to change the current circumstances
- **Positive reinterpretation:** increase utility of positive side-effect of an act with a negative outcome
- **Acceptance:** drop a threatened intention
- **Denial:** lower the probability of a pending undesirable outcome
- **Mental disengagement:** lower utility of desired state
- **Shift blame:** shift responsibility for an action toward some other agent
- **Seek/suppress information:** form a positive or negative intention to monitor some pending or unknown state

Strategies give input to the cognitive processes that actually execute these directives. For example, planful coping will generate an intention to perform the “join club” action, which in turn leads the planning system to generate and execute a valid plan to accomplish this act. Alternatively, coping strategies might abandon the goal, lower the goal’s importance, or re-assess who is to blame.

Not every strategy applies to a given stressor (e.g., an agent cannot engage in problem directed coping if it is unaware of an action that impacts the situation), however multiple strategies can apply. EMA proposes these in parallel but adopts strategies sequentially. EMA adopts a small set of search control rules to resolve ties. In particular, EMA prefers problem-directed strategies if control is appraised as high (take action, plan, seek information), procrastination if changeability is high, and emotion-focus strategies if control and changeability is low.



Figure 3: The MRE and SASO-ST systems allow a trainee to interact with intelligent virtual characters through natural language for task-oriented training.

In developing a computational model of coping, we have moved away from the broad distinctions of problem-focused and emotion-focused strategies. Formally representing coping requires a certain crispness lacking from the problem-focused/emotion-focused distinction. In particular, much of what counts as problem-focused coping in the clinical literature is really inner-directed in an emotion-focused sense. 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 one's interpretation of circumstances without actually changing the physical environment.

Appraisal theory and the design of virtual humans

The question we raise in this chapter is the connection between emotion research and the design of cognitive systems. We have explored this question within our own work in the context of the design of virtual humans. These are software agents that attempt to simulate human cognitive, verbal and nonverbal behavior in interactive virtual environments. From the perspective of this volume, virtual humans serve to illustrate the complexity of contemporary cognitive systems and the host of integration and control problems they raise. After describing the capabilities of such agents, we show how our understanding of human emotion, and computational appraisal theory in particular, has influenced the design of a general architecture that can detect, classify, and adaptively respond to significant changes to their virtual environment.

Figure 3 illustrates two applications of this architecture that support face-to-face multi-modal communication between users and virtual characters in the context of interpersonal-skills. In the Mission Rehearsal Exercise (MRE), the learner plays the role of a lieutenant in the U.S. Army involved in a peacekeeping operation in Bosnia (Swartout et al., 2001). In route to assisting another unit, one of the lieutenant's vehicles becomes involved in a traffic accident, critically injuring a young boy, the boy's mother is understandably distraught and a local crowd begins to gather. The learner must resolve the situation by interacting through spoken language with virtual humans in the scene, and learn to juggle multiple, interacting goals (i.e., assisting the victim vs. continuing his mission). In the Stability and Support Operations – Simulation and Training (SASO-ST) exercise, the learner plays the role of a captain assisting security and reconstruction

efforts in Iraq (Traum, Swartout, Marsella, & Gratch, 2005) and must negotiate with a simulated doctor working with a non-governmental aid organization and convince him to move his clinic to another location. The learner must resolve the situation by interacting through spoken language and by applying principles of effective negotiation. In both applications, agents must react in real time to user dialogue moves, responding in a way that is appropriate to the agent's goals and subject to the exercise's social and physical constraints.

An Integration Challenge

The AUSTIN virtual human architecture underlying these applications must integrate a diverse array of capabilities.³ Virtual humans develop plans, act and react in their simulated environment, requiring the integration of automated reasoning and planning techniques. To hold a conversation, they demand the full gamut of natural language research, from speech recognition and natural language understanding to natural language generation and speech synthesis. To control their graphical bodies, they incorporate real-time graphics and animation techniques. And because human movement conveys meaning, virtual humans draw heavily on psychology and communication theory to appropriately convey nonverbal behavior. More specifically, AUSTIN integrates:

- a task reasoning module that allows virtual humans to develop, execute and repair team plans and to reason about how past events, present circumstances, and future possibilities impact individual and team goals. Agents use domain-independent planning techniques incorporating elements of decision-theoretic plan representations with explicit representations of beliefs, intentions and authority relationships between individuals. (Rickel et al., 2002; Traum, Rickel, Gratch, & Marsella, 2003), and must balance multiple goals and multiple alternative plans for achieving them.
- a realistic model of human auditory and visual perception (Kim, Hill, & Traum, 2005) that restricts perceptual updates to information what is observable, given the constraints of the physical environment and character's field of view. Although this has the benefit of reducing perceptual processing and renders the virtual human's behavior more realistic, limited perception introduces the control problem of what features of the environment should be actively attended.
- a speech understanding module that incorporates a finite-state speech recognizer and a semantic parser to produces a semantic representation of utterances (Feng & Hovy, 2005). These interpretations may be underspecified, leading to perceptual ambiguity in the speech processing that raises a host of control decisions (e.g., should the agent clarify the ambiguity or should it choose the most likely interpretation).
- a dialogue model that explicitly represents aspects of the social context (Matheson, Poesio, & Traum, 2000; Traum, 1994) while supporting multi-party conversations and face-to-face communication (Traum & Rickel, 2002). This module must make a variety of choices in concert with other action selection decisions in the agent: It must choose amongst a variety of speech acts including dialogue acts that can influence who has the conversational turn, what topic is under discussion, whether to clarify or assume, etc.
- a natural language generator that must assemble and choose between alternative utterances to convey the agent's speech act. This can produce nuanced English expressions

that vary depending on the virtual human's emotional state as well as the selected content (Fleischman & Hovy, 2002)

- an expressive speech synthesizer capable of choosing between different voice modes depending on factors such as proximity (speaking vs. shouting) and illocutionary force (command vs. normal speech) (Johnson et al., 2002)
- a gesture planner that that assembles and chooses between alternative nonverbal behaviors (e.g., gestures, head movements, eyebrow lifts) to associate with the speech (Marsella, Gratch, & Rickel, 2003). This module augments the BEAT system (Cassell, Vilhjálmsón, & Bickmore, 2001) to incorporate information about emotional state as well as the syntactic, semantic and pragmatic structure of the utterance.
- a procedural animation system developed in collaboration with Boston Dynamics, Inc. supports the animation and rendering of the virtual character.
- a control system, based on appraisal and coping, that characterizes the current task and dialogue state in terms of appraisal variables, and suggests a strategic response that informs choices made by other modules including the perception module, task reasoner, dialogue manager, language generator and gesture planner.

This integration raises serious control and coordination problems are similar to the issues emotion is posited to address. The agent must divide cognitive resources between plan generation, monitoring features of the environment, and attending to a conversation. But because the agent is embodied with a human-like appearance and communicates through naturalistic methods, this becomes far more than a traditional scheduling problem. For example, if an agent takes several seconds to respond to a simple yes-or-no questions, users will become annoyed or read too much into the delay (one trainee felt the character was angry with them as a result of a bug that increased dialogue latency). Further, the agent must maintain some sense of consistency across its various behavioral components including the agent's internal state (for example, goals, plans, and emotions) and the various channels of outward behavior (for example, speech and body movements). When real people present multiple behavior channels, observers interpret them for consistency, honesty, and sincerity, and for social roles, relationships, power, and intention. When these channels conflict, the agent might simply look clumsy or awkward, but it could appear insincere, confused, conflicted, emotionally detached, repetitious, or simply fake.

This cognitive architecture builds on prior work in the areas of embodied conversational agents (Cassell, Sullivan, Prevost, & Churchill, 2000) and animated pedagogical agents (Johnson, Rickel, & Lester, 2000), but integrates a broader set of capabilities than such systems. Classic work on virtual humans in the computer graphics community focuses on perception and action in 3D worlds (Badler, Phillips, & Webber, 1993; Thalmann, 1993), but largely ignores dialogue and emotions. Several systems have carefully modeled the interplay between speech and nonverbal behavior in face-to-face dialogue (Cassell, Bickmore, Campbell, Vilhjálmsón, & Yan, 2000; Pelachaud, Badler, & Steedman, 1996) but these virtual humans do not include emotions and can not participate in physical tasks in 3D worlds. Some work has begun to explore the integration of conversational capabilities with emotions (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000; Marsella, Johnson, & LaBore, 2000; Poggi & Pelachaud, 2000), but still does not address physical tasks in 3D worlds. Likewise, prior work on STEVE addressed the issues of integrating face-to-face dialogue with collaboration on physical tasks in a 3D virtual world (Rickel & John-

son, 2000), but STEVE did not include emotions and had far less sophisticated dialogue capabilities than our current virtual humans. The tight integration of all these capabilities is one of the most novel aspects of our current work. The AUSTIN cognitive architecture seeks to advance the state of the art in each of these areas, but also to explore how best to integrate them into a single agent architecture, incorporating a flexible blackboard architecture to facilitate experiments with the connections between the individual components.

Emotion, Design and Control

We claim that appraisal theory provides a unifying conceptual framework that can inform the design of complex cognitive systems. We illustrate how it has informed our approach to integration and control of the AUSTIN cognitive architecture. The control and integration issues arising from the AUSTIN are hardly unique to virtual humans. The problem of allocating computational resources across diverse functions, coordinating their activities and integrating their results is common to any complex system. The solutions to such problems, however, have tended to be piecemeal as research has tended to focus on a specific control issue, for example “exploration vs. exploitation” or “planning vs. acting.” In contrast, we argue that appraisal theory provides a single coherent perspective for conceptualizing cognitive control.

Adopting an appraisal theoretic perspective translates into several proscriptions for the design of a cognitive system.

Appraisal as a uniform control structure: Appraisal theory suggests a general set of criteria and control strategies that could be uniformly applied to characterize, inform and coordinate the behavior of heterogeneous cognitive functions. Whether it is processing perceptual input or exploring alternative plans, cognitive processes must make similar determinations: is the situation/input they are processing desirable and expected; does the module have the resources to cope with its implications? Such homogenous characterizations are often possible, even if individual components differ markedly. By casting the state of each module in these same general terms, it becomes possible to craft general control strategies that apply across modules.

Further, appraisal theory argues that each appraisal variable provides critical information that informs the most adaptive response. For example, if there is a threat on the horizon that may vanish of its own accord, it is probably not worth cognitive resources to devise a contingency and an organism should procrastinate; if the threat is looming and certain, an organism must act and its response should vary depending on its perceived sense of control: approach (i.e., recruit cognitive or social resources to confront the problem) if control is high; avoid (i.e., retreat from the stressor or abandon a goal) if control is low. From an ecological perspective (see [Todd and Kirlik in this volume](#)), these mappings can be viewed as simple control heuristics that suggest appropriate guidance for the situations an organism commonly experiences, and may translate into robust control strategies for cognitive systems.

In AUSTIN, we have explored this principle of control uniformity to the design of two core components, the plan-reasoning module and the dialogue manager. Beside the plan-based appraisal and coping described above, AUSTIN introduces analogous techniques to characterize the current state of a dialogue in terms of appraisal variables (e.g., what is the desirability of a particular dialogue tactic, how likely it is to succeed and how much control an agent has over this success),

and crafted alternative dialogue strategies that mirror the plan and emotion-focused coping strategies available to the planning system.

Besides simplifying AUSTIN's control architecture, this principle offered insight on how to elegantly model and select amongst alternative dialogue strategies. For example, the SASO-ST system is designed to teach principles of negotiation, including the competitive/cooperative orientation of the parties to the negotiation and the strategies they employ in light of those orientations. Specifically, one oft-made distinction is between integrative vs. distributive stances towards negotiation (Walton & Mckersie, 1965). A distributive stance occurs when parties interpret a negotiation as zero-sum game where some fixed resource must be divided, whereas an integrative stance arises when parties view the situation as having mutual benefit. Thirdly, parties may simply believe that there is no possible benefit to the negotiation and simply avoid the negotiation or deny the need for it, what is termed avoidance (e.g., (Sillars, Coletti, Parry, & Rogers, 1982)). Although described with different terminology, there are strong conceptual similarities between this theory of negotiation and appraisal theory: both argue that response strategies are influenced by an appraisal of the current situation. For example, if the outcome of a negotiation seems undesirable but avoidable, the agent adopts a strategy to disengage (e.g., change topics). If these attempts fail, the agent may re-appraise the situation as less controllable and thus more threatening, motivating distributive strategies. By adopting an appraisal-theoretic perspective, we are able to recast negotiation stances as alternative strategies for coping with the appraised state of the negotiation, and thereby leverage the existing appraisal/coping machinery.

Appraisal as a value computation: Appraisal can be seen as a utility calculation in the sense of decision theory, and thus can subsume the role played by decision theory in cognitive systems. For example, it can determine the salience and relative importance of stimuli. The difference is that appraisal can be seen as a multi-attribute function that incorporates broader notions than simply probability and utility. In particular, it emphasizes the importance of control – does the agent have the power to affect change over the event – which, according to appraisal theory, is critical for determining response. Thus, appraisal theory can support the value computations presumed by many mental functions, but support subtler distinctions than traditional cognitive systems.

In AUSTIN, appraisal acts as a common currency for communicating the significance of events between the planning, dialogue management, and perceptual modules and facilitates their integration. One example of this is determining linguistic focus. In natural language, people often speak in imprecise ways and one needs to understand the main subject of discussion to correctly disambiguate meaning. For example, when the trainee encounters the accident scene in the MRE scenario, he might ask the virtual human, “What happened here?” In principle many things have happened: the trainee just arrived, the soldiers assembled at the meeting point, an accident occurred, a crowd formed, and so forth. The virtual human could talk about any one of these and be factually correct, but not necessarily pragmatically appropriate. Rather, people are often focused most strongly on the things that upset them emotionally, which suggests an emotion-based heuristic for determining linguistic focus. Because we model the virtual character's emotions, the dialogue planning modules have access to the fact that he is upset about the accident can use that information to give the most appropriate answer: describing the accident and how it occurred.

Another example is the integration of top-down and bottom-up attention in the control of perception. In AUSTIN, the virtual human must orient its sensors (virtual eyes) to stimuli in order to perceive certain changes in the environment, which raises the control problem of what to look at next. This decision can be informed by bottom up processes that detect changes in the environment (e.g. (Itti & Koch, 2001)) and by top-down processes that calculate the need for certain information. We have been exploring the use of appraisal as a value calculation to inform such top down processes. Thus, for example, attention should be directed towards stimuli generating intense appraisals.

Appraisal as a design specification for cognition: Appraisal theory presumes that an organism can interpret situations in terms of several criteria (i.e., appraisal variables) and use this characterization to alter subsequent cognitive processing (e.g. approach, avoidance or procrastination). On the one hand, these assumptions dictate what sort of inferences a cognitive system must support. On the other hand, they argue that inferential mechanisms must support qualitatively different processing strategies, sensitive to the way input is appraised. Traditional cognitive systems consider only a subset of these criteria and strategic responses. In terms of appraisal, for example, cognitive systems do a good job about reasoning about an event's desirability and likelihood, but rarely consider the social factors that inform causal attributions. In terms of coping, cognitive systems excel at problem-focused strategies (e.g., planning, acting, seeking instrumental social support) but have traditionally avoided emotion-focused strategies such as goal abandonment and denial.

Adopting this perspective, we identified several missing capabilities in the AUSTIN cognitive architecture, particularly as it relates to human social behavior. In its early incarnation, for example, AUSTIN used physical causality as a proxy for human social inference. In terms of the appraisal variable of causal attribution, this translates into the inference that if a person performed an action with some consequence, they deserve blame for that consequence. However, appraisal theory identifies several critical factors that mediate judgments of blame and responsibility for social activities, including whether the person intended the act, were aware of the consequence and if their freedom to act was constrained by other social actors. Before making such inferences, AUSTIN would make inappropriate attributions of blame, such as blaming individuals when their actions were clearly coerced by another agent. Subsequent research has illustrated how to incorporate such richer social judgments into the architecture (Mao & Gratch, 2005).

This principle also led to the modeling of emotion-focused coping strategies, important for increasing the cognitive realism of the agent but also of potential value for managing commitments and cognitive focus of attention. Following Pollack (Pollack, 1990), commitments to goals and beliefs can be viewed as control heuristics that prevent the expenditure of cognitive resources on activities inconsistent with these commitments. 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. Appraisal theory suggests a novel solution to the problem of when to abandon commitments that we have incorporated into AUSTIN. 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 (Mele, 2001). This suggests that there is no simple criterion for abandoning commitments, but rather one must weight the pros and cons of alternative con-

flicting commitments. Appraisal and coping provide a possible mechanism for providing this evaluation. Appraisal identifies particularly strong conflicts in the causal interpretation, whereas coping assesses alternative strategies for resolving the conflict, dropping one conflicting intention or changing some belief so that the conflict is resolved.

Conclusion

As cognitive systems research moves beyond simple, static and nonsocial problem solving, researchers must increasingly confront the challenge of how to allocate and focus mental resources in the face of competing goals, disparate and asynchronous mental functions, and events that unfold across a variety of time scales. Human emotion clearly exerts a controlling influence over cognition and here we have argued that a functional analysis of emotion's impact can profitably inform the control of integrated cognitive systems. Computational appraisal theory, in particular, can help translate psychological findings about the function of emotion into concrete principles for the design of cognitive systems. Appraisal theory can serve as a blueprint for designing a uniform control mechanism for disparate cognitive functions, suggesting that the processing of these individual components can be uniformly characterized in terms of appraisal variables, and controlled through a common mapping between appraisal and action tendency (coping). Appraising the activities of individual components also allows emotion to act as a common currency for assessing the significance of events on an agent's cognitive activities. Finally, as a theory designed to characterize emotional responses to a wide span of human situations, appraisal theory can serve as a requirements specification, suggesting core cognitive functions often overlooked by traditional cognitive systems. These principles have influenced the course of our own work in creating interactive virtual humans and, we contend, can profitably contribute to the design of integrated cognitive systems.

Acknowledgements

We gratefully acknowledge the feedback of Wayne Gray and Mike Schoelles on an earlier draft of this paper. This work was sponsored by the U. S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

- Allmana, J., Hakeema, A., Erwinb, J., Nimchinskyc, E., & Hofd, P. (2001). The Anterior Cingulate Cortex: The Evolution of an Interface between Emotion and Cognition. *Annals of the New York Academy of Sciences*, 935, 107-117.
- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Arnold, M. (1960). *Emotion and Personality*. NY: Columbia University Press.
- Badler, N. I., Phillips, C. B., & Webber, B. L. (1993). *Simulating Humans*. New York: Oxford University Press.
- Bates, J., Loyall, B., & Reilly, W. S. N. (1991). Broad Agents. *Sigart Bulletin*, 2(4), 38-40.
- Bechara, A., Damasio, H., AntonioDamasio, & Lee, G. (1999). Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making. *Journal of Neuroscience*, 19(13), 5473-5481.

- Bless, H., Schwarz, N., & Kimmelmeier, M. (1996). Mood and stereotyping: The impact of moods on the use of general knowledge structures. *European Review of Social Psychology*, 7, 63-93.
- Blythe, J. (1999, Summer). Decision Theoretic Planning. *AI Magazine*, 20(2), 37-54.
- Bower, G. H. (1991). Emotional mood and memory. *American Psychologist*, 31, 129-148.
- Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsón, 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., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
- Cassell, J., Vilhjálmsón, H., & Bickmore, T. (2001). *BEAT: The Behavior Expressive Animation Toolkit*. Paper presented at the SIGGRAPH, Los Angeles, CA.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.
- 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.
- Feng, D., & Hovy, E. H. (2005). *MRE: A Study on Evolutionary Language Understanding*. Paper presented at the Proceedings of the Second International Workshop on Natural Language Understanding and Cognitive Science (NLUCS), Miami, FL.
- 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, NY: W. W. Norton.
- Frijda, N. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, 1, 115-143.
- 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. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.
- Gratch, J., & Marsella, S. (2005). Evaluating a computational model of emotion. *Journal of Autonomous Agents and Multiagent Systems* (in press).
- Grosz, B., & Kraus, S. (1996). Collaborative Plans for Complex Group Action. *Artificial Intelligence*, 86(2), 269-357.
- Itti, L., & Koch, C. (2001). Computational Modeling of Visual Attention. *Nature Reviews Neuroscience*, 2(3), 194-203.
- 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.
- Johnson, W. L., Rickel, J., & Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of AI in Education*, 11, 47-78.

- Kim, Y., Hill, R. W., & Traum, D. R. (2005). *A Computational Model of Dynamic Perceptual Attention for Virtual Humans*. Paper presented at the Proceedings of 14th Conference on Behavior Representation in Modeling and Simulation (BRIMS), Universal City, CA.
- Lazarus, R. (1991). *Emotion and Adaptation*. NY: Oxford University Press.
- 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: MIT Press.
- Mao, W., & Gratch, J. (2004). *Social Judgment in Multiagent Interactions*. Paper presented at the Third International Joint Conference on Autonomous Agents and Multiagent Systems.
- Mao, W., & Gratch, J. (2005). *Social Causality and Responsibility: Modeling and Evaluation*. Paper presented at the International Working Conference on Intelligent Virtual Agents, Kos, Greece.
- 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 Multi-agent 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* (pp. 317-360). Berlin: Springer-Verlag.
- Marsella, S., Johnson, W. L., & LaBore, C. (2000). *Interactive Pedagogical Drama*. Paper presented at the Fourth International Conference on Autonomous Agents, Montreal, Canada.
- Matheson, C., Poesio, M., & Traum, D. (2000). *Modeling Grounding and Discourse Obligations Using Update Rules*. Paper presented at the First Conference of the North American Chapter of the Association for Computational Linguistics.
- Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.
- Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*: Cambridge University Press.
- Peacock, E., & Wong, P. (1990). The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal. *Stress Medicine*, 6, 227-236.
- Pelachaud, C., Badler, N. I., & Steedman, M. (1996). Generating Facial Expressions for Speech. *Cognitive Science*, 20(1).
- 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.
- Rickel, J., & Johnson, W. L. (2000). Task-Oriented Collaboration with Embodied Agents in Virtual Worlds. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents*. Boston: MIT Press.
- 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.
- Russell, J. A., & Lemay, G. (2000). Emotion Concepts. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 491-503). New York/London: Guilford Press.

- 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).
- Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology*, 24, 161-199.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. NY: Springer-Verlag.
- Sillars, A. L., Coletti, S. F., Parry, D., & Rogers, M. A. (1982). Coding verbal conflict tactics: Nonverbal and perceptual correlates of the avoidance-distributive- integrative distinction. *Human Communication Research*, 9, 83-95.
- Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, 74, 29-39.
- Smith, C. A., & Lazarus, R. (1990). Emotion and Adaptation. In L. A. Pervin (Ed.), *Handbook of Personality: theory & research* (pp. 609-637). NY: Guilford Press.
- Swartout, W., Hill, R., Gratch, J., Johnson, W. L., Kyriakakis, C., LaBore, C., et al. (2001). *Toward the Holodeck: Integrating graphics, sound, character and story*. Paper presented at the Fifth International Conference on Autonomous Agents, Montreal, Canada.
- Thalmann, D. (1993). Human Modeling and Animation. In *Eurographics '93 State-of-the-Art Reports*.
- Traum, D. (1994). *A computational theory of grounding in natural language conversation*. Unpublished Ph.D., University of Rochester, Rochester, NY.
- 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. (2003). *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.
- Traum, D., Swartout, W., Marsella, S., & Gratch, J. (2005). *Fight, Flight, or Negotiate*. Paper presented at the Intelligent Virtual Agents, Kos, Greece.
- Walton, R. E., & Mckersie, R. B. (1965). *A behavioral theory of labor negotiations: An analysis of a social interaction system*. McGraw-Hill.
- Weiner, B. (1995). *The Judgment of Responsibility*: Guilford Press.

¹ “Appraisal Theory” is commonly used to refer to a collection of theories of emotion that agree in their basic commitments but vary in detail and process assumptions. Here we emphasize their similarity. See Ellsworth and Scherer for a discussion the similarity and differences between competing strands of the theory (Ellsworth & Scherer, 2003) In our own work, we are most influenced by the conception of appraisal theory advocated by Richard Lazarus.

² EMA stands for EMotion and Adaptation, the title of the book by Richard Lazarus that most influenced the development of the model.

³ AUSTIN is an incremental extension of our earlier STEVE system.