

DOI:10.1145/2631912

**Computer models of emotion inform theories of human intelligence and advance human-centric applications.**

BY STACY MARSELLA AND JONATHAN GRATCH

# Computationally Modeling Human Emotion

EMOTION'S ROLE IN human behavior is an old debate that has become increasingly relevant to the computational sciences. Two-and-a-half millennia ago, Aristotle espoused a view of emotion at times remarkably similar to modern psychological theories, arguing that emotions (such as anger), in moderation, play a useful role, especially in interactions with others. Those who express anger at appropriate times are praiseworthy, while those lacking in anger at appropriate times are treated as a fool. The Stoics took a different view; four centuries after Aristotle, Seneca considered emotions (such as anger) as a threat to reason, arguing, "reason ... is only powerful so long as it remains isolated from emotions." In the 18<sup>th</sup> century, David Hume radically departed from the Stoic perspective, arguing for the key motivating role of emotions, saying, "Reason is, and ought only to be the slave of the passions."

A similar dichotomy of views can be seen in the history of artificial intelligence (AI) and agent research.

Early work by Herbert A. Simon<sup>35</sup> argued that emotions served a critical function in intelligent behavior, as an interrupt capacity that provides a means for an organism to shift between competing goals, as well as to balance reactive and deliberative processing. Marvin Minsky posed his question of whether a robot could even be intelligent without emotion. However, late-20<sup>th</sup> century AI research took a more Stoic perspective, treating emotion as antithetical to intelligence.

Modern research in psychology and neuroscience continues to transform the debate. Appraisal theories of emotion<sup>21</sup> emphasize the role of reasoning in eliciting emotion, viewing emotion as arising from people's "appraisal" of their relationship to the environment that guides adaptive responses. Recent studies establish emotion's critical role in decision making, a point highlighted by the fact that neurological deficits in emotion processing lead to deficits in decision making.<sup>3</sup> In keeping with Simon's view of emotion as interrupt, emotions prime perceptions and response patterns. Angry people are, for example, quicker to perceive threats<sup>11</sup> and are generally primed for an aggressive response.<sup>20</sup>

Echoing Aristotle's view, research has also argued emotions and their expression play a powerful, adaptive role in social interaction; for example, emotional displays convey informa-

## >> key insights

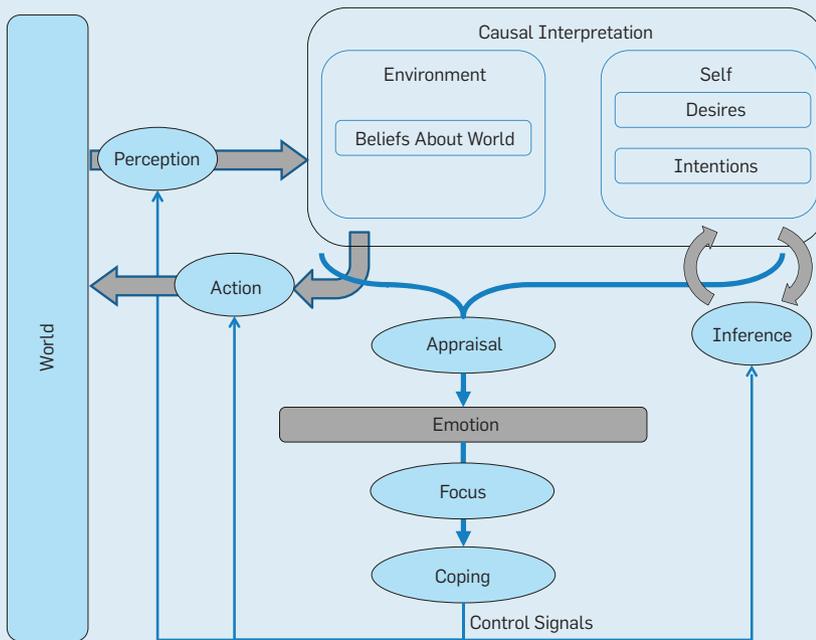
- Processes akin to emotion are required by any intelligent entity facing a dynamic, uncertain, and social environment.
- Psychological theories of emotion (such as appraisal theory) can serve as an architectural specification for machines that aim to recognize, model, and simulate human affect.
- Realizing psychological theories as working computational models advances science by forcing concreteness, revealing hidden assumptions, and creating dynamic artifacts that can be subject to empirical study.



**Figure 1. Illustration of the dynamics of emotional reactions.**



**Figure 2. EMA's appraisal, coping, and reappraisal.**



tion about an individual's beliefs, desires, and intentions and thereby serve to inform and influence others. Anger and guilt can improve group utility by minimizing social conflicts,<sup>15</sup> while an expression of distress induces social support.<sup>13</sup> What makes these signals powerful, as Hume suggested, is the shared knowledge of the motivational power of emotion. We may believe someone feels a situation unpleasant, but, often, what motivates us to act is the emotional content perceived in the person's behavior, a parent's anger, or a child's cries of distress.

This research has spawned a rebirth of interest in modeling emotion in AI, robotics, and agent research. In particular, software agent and robotics research has explored computational models of emotion as a way to address control and decision trade-

offs by directing cognitive resources toward problems of adaptive significance for the organism.<sup>5,7,33,36</sup>

Research in autonomous agents and multi-agent systems, as well as in human-computer interaction (such as Conati and MacLaren<sup>8</sup>), have explored how to exploit emotion's social function in facilitating interactions between computer systems and human users. Emotions and emotional displays have been used in virtual characters to exploit the social function of emotional expression to motivate<sup>22</sup> and establish empathy and bonding.<sup>27</sup>

Research in computational models also influences theories of human emotion by transforming how these theories are formulated and evaluated. Realizing a theory as a computational model requires processes and their interactions to be formally detailed,

thereby exposing implicit assumptions and hidden complexities. Incorporating the models into larger simulations further exposes hidden issues and can extend the scope of the theory; for example, incorporating emotion into more-comprehensive simulations of human behavior has led researchers to address fundamental questions about the relation of appraisal processes to other cognitive processes, perception, and behavior.<sup>4,6,12,25</sup>

On a conceptual level, computational models enrich the language-of-emotion theories; for example, several computational models have recast emotion theories in terms of concepts drawn from AI, including knowledge representation (such as Gratch and Marsella<sup>17</sup>), planning (such as Dias and Paiva<sup>12</sup>), and neural networks (such as Armony et al.<sup>1</sup>).

On an empirical level, computational models facilitate an expanded range of predictions compared to conventional theory. Computer simulations are a way to explore the temporal dynamics of emotion processes and form predictions about the source and time course of those dynamics. Manipulations of experimental conditions may be explored more extensively first with a computational model, as in ablating functionalities or testing responses under adverse conditions that may be costly, risky, or raise ethical concerns in vivo.<sup>1</sup> Simulations can reveal unexpected model properties that suggest further exploration. Additionally, models of emotion and affective expression have been incorporated into virtual humans,<sup>37</sup> software artifacts that look and act like humans, capable of interacting with people in a virtual world they cohabit. These systems essentially allow for the study of emotion in a virtual ecology.

Here, we discuss our work on a

computational model of emotion, detailing the design principles that drove the work, as well as its realization within a cognitive architecture. As we envision, a fundamental challenge for a computational model of emotion is how to address how emotions arise and evolve over a range of eliciting situations, from physical stimuli to complex social situations, in keeping with roles in individual behavior and social interaction. These emotional responses can be quick and reactive or seemingly more deliberative, unfolding over minutes, days, even weeks; for example, anger at a co-worker's behavior may not arise immediately but rather require inferences about the underlying causes of the behavior. In short, emotions are inherently dynamic, linked to both the world's dynamics and the dynamics of the individual's physiological, cognitive, and behavioral processes.

Drawing on a leading psychological theory of emotion, we discuss how a computational model of emotion processes can address both the time course of emotions and the range of eliciting conditions. After laying out the model, we discuss how it has transformed our thinking on emotion and the old debate on the relation of cognition to emotion. We then discuss an example use of the model in AI and address validation. Whether one is using a model of emotion to improve some application or as a methodological tool in the study of human emotion, a central question is how to validate the model in light of such uses. First, however, we discuss a salient example of emotion's role in behavior.

### Example

To help ground our discussion of appraisal theory and our computational approach to modeling it, we describe a real-world emotion-evoking situation. Recorded by happenstance, it serves to illustrate both the rich dynamics of emotion processes and the range of eliciting conditions that induce emotions. As part of our work on modeling virtual humans, we were doing an improvisational session with actors to record their use of nonverbal behavior. As the rehearsal wore on into a hot summer Los Angeles night, a dove flew into an open window; Figure 1 covers

the sequence of reactions of one of the actors. Although such an uncontrolled event makes rigorous analysis of reactions uncertain, we suggest the following interpretation:

The bird flew into the window and then hit a windowsill in a failed attempt to fly back out. The actor proceeded to orient toward the stimulus of the sound. The first reaction was apparent surprise, revealed by the one of the features of surprise discussed by Charles Darwin<sup>9</sup> involving raised eyebrows, serving to both increase the field of view and warn others of an unexpected event.

The eyebrows then lowered, and the mouth opened. Mouth opening can have functional significance, as a visual or audible signal (such as shouting) or to draw in air to oxygenate the blood in preparation for a fight or flight response. The lowered eyebrows suggested a negative affective response (fear).

At this point the actor moved away from the threat of the dove (a flight response) but at the same time started to assume a more aggressive, defensive stance. Her hands moved on the umbrella in preparation of using it as a weapon to possibly whack the bird (a fight response).

She continued to quickly move away (approximately seven feet), and in the process the umbrella was lowered and a hand went up to her mouth, as she reacted to others' predicament in the room. Her hand quickly went into a classic stop (or be careful) gesture as she reversed direction and started to move toward the bird. Her concern now shifted from herself to the bird, as the bird was caught in the hair of a fellow actor and in danger of being hurt. In the final frame in Figure 1, the actor has a handkerchief she will use to rescue the bird by wrapping and releasing it without harm out the window. However, after the bird was safely released out the window, the actors' mood remained in a heightened state that made continued rehearsal impossible.

We see in this interpretation of events a variety of functional roles for emotion relevant to researchers in AI, autonomous agents, and multi-agent systems.

Clearly illustrated is emotion's function in interrupting and re-prioritizing cognition, as discussed by Simon.<sup>35</sup> We

see interruption of current attention, cognitive processes, and physical responses that shift and recruit resources to deal with a new threatening situation, specifically to gather more information relevant to one's self, interrupt current goals associated with the improvisation session, and switch to the goal of dealing with a potential threat, taking action in preparation to dealing with that threat at both the physical level (such as fight and flight) and physiological responses (such as sucking in oxygen to oxygenate the blood). All this was in service to adjusting to a dynamic relation between the actor and her environment. The situation was changing, due, in part, to events external to the actor (such as a bird flying in the window toward the actor), as well as to the actor's own response (such as "arming oneself" and moving away from the event). The actor's interpretation of the situation also appeared to evolve—bird as threat versus bird as victim—suggesting perceptual and inferential processes with their own internal dynamics, requiring time to draw inferences and reassess the situation and replan, as new knowledge is brought to bear. There were more persistent dynamics as well. The altered mood that persisted after the event could play a functional role. Such heightened arousal can prepare an individual for other quick responses, assuming the event presages future similar occurrences.

As events unfolded, emotion's function in improving multi-agent coordination and group utility<sup>15</sup> was also clearly on display. In particular, emotion and its expression play a role in signaling mental states, including emotions and intentions that influence interaction; for example, expressions of anger may ward off the hostile agent. The expressions of emotion can also help orient and coordinate group response; the expression of fear that signals a common threat to members of the agent's own group and expression of compassion that can influence interpersonal attitudes and empathy, thereby serving to establish common goals.

We also see in this quick evolution of reactions and behaviors the powerful role of emotions as motivator, echoing Hume's view of the motivating role of emotions.

Finally, the trajectory of events in the rehearsal studio also reveals emotion's rapidly changing nature. Overall, the observed reactions suggest a progression from surprise about the unexpected event, concern for protecting self, and finally concern for others, including the original source of the perceived threat—the bird. This happened quite rapidly. Within approximately 2.6 seconds, the goals went from flight to fight to helping the bird. The expression of raised eyebrows often associated with surprise took approximately 60 milliseconds and the expression of lowered eyebrows and lowered jaw (often associated with anger and responses to threat) approximately 300 milliseconds. Tightly coupled with these evolving concerns from threat-to-self to threat-to-other and the emotion dynamics of fear/anger to compassion/relief is a corresponding progression of coping responses from defend/attack to help.

As interpreted here, this example presents fundamental challenges for a model of emotion processes. First, a model must be able to address how emotions arise and evolve over a range of eliciting conditions, from simple physical events to complex social situations. Second, emotional responses can be rapid, on the order of milliseconds, or unfold over minutes or days, following some delib-

eration or inference. Emotion's dynamics are linked to both the world's dynamics and the dynamics of the individual's physiological, cognitive, and behavioral processes.

### From Theory to Model

Computational models of emotion often rely on psychological theories of emotion as a basis for a computational model. Minimally, the theory serves to define what an emotion is, as in what comprises an emotional state like anger. Also, theories may lay out the antecedents and consequents of an emotion, factors critical to development of a computational model of the emotion process.

Complicating development of the computational model of emotion is that there are very different psychological theories of emotion. Theories differ as to which components are treated as intrinsic to an emotion (such as cognitions, somatic processes, behavioral tendencies, and responses), the relationships between components (such as whether cognitions precede or follow somatic processes), and representational distinctions (such as whether anger is a linguistic fiction or a natural kind).

Describing the range of theories that can inform model development is beyond our scope here; for a more complete discussion, see Marsella et al.<sup>26</sup> Nevertheless, we can very broadly

characterize many modern theories into three categories:

*Discrete theories of emotion.*<sup>14</sup> These theories argue there is a limited number of core emotions that are biologically determined, innate natural kinds (such as anger and sadness). Further, their expression is shared across people and cultures. A related view conceptualizes emotions as distinct neural circuits;<sup>29</sup> see Armony<sup>1</sup> for an example of a theory of conditioned fear response and associated neural network model.

*Dimensional theories of emotion.* These theories argue that emotion and other affective phenomena should be conceptualized, not as discrete categories but as a point in a continuous (typically 2D or 3D) space.<sup>2,28</sup> Dimensional theories often argue discrete emotion categories (such as anger) do not have a specific biological basis; that is, there is no brain region or circuit that is unique to that emotion category. Rather, they are folk-psychological labels people assign to loosely coupled collections of mental and physical states.<sup>2</sup> Computational models that build on dimensional theories often use the “PAD” theory of Mehrabian and Russell,<sup>28</sup> its three dimensions corresponding to pleasure (a measure of valence), arousal (indicating the level of affective activation), and dominance (a measure of power or control). Several computational models rely on dimensional theories, to, say, control action selection, model an embodied feeling state that is mapped to emotion categories, and model moods.<sup>4,7,16</sup>

*Appraisal theories of emotion.* These theories arose from attempts to detail the mental processes underlying the elicitation of emotions.<sup>21,32</sup> Just as the actor in our example had to reconcile her goals and needs in relation to a rapidly changing environment, appraisal theories say emotions arise from a process of comparing individual needs to external demands. That is, emotions cannot be explained by solely focusing on the environment or by solely focusing on the individual. Rather, they reflect the “person-environment relationship.” Appraisal theories further posit this person-environment relationship is characterized (or appraised) in terms of a set of criteria, variously called appraisal variables, checks, or



dimensions; for example, Is this event desirable with respect to one's goals or concerns?; Who caused it?; Was it expected?; and What power do I have over its unfolding? The results of these appraisal checks are in turn mapped to emotion. Some specific variants of appraisal theories then go on to detail how the resulting emotions influence the individual's cognitive and behavioral responses; for example, Lazarus<sup>21</sup> detailed how the emotions lead to coping responses that seek to change the world (problem-directed coping) or one's own emotional response (emotion-directed coping).

Beyond its impact in psychology, appraisal theories have become the dominant framework for building computational models of emotion; see Reisenzein et al.<sup>31</sup> for more on models that use appraisal theories. This dominance reflects, in part, its prominence in emotion psychology. More fundamentally, however, the concept of appraisal (involving judgments like desire, expectedness, and cause) maps nicely to traditional AI concepts (such as the belief-desire-intention models of agency), a point we emphasize when describing our appraisal-based model.

Appraisal theories are not without criticism. Indeed, the characteristic that makes them most attractive to computational modelers—emphasis on the role of inferential processes—is also its most controversial feature within psychology. By emphasizing judgments over some explicit representation of a person-environment relationship, appraisal theories are traditionally viewed as requiring expensive cognitive inference as a precursor to emotional reactions; that is, an individual must engage in inference before responding emotionally to some stimuli. This characteristic seems to fly in the face of the rapid and seemingly automatic emotional reactions we see in, say, our actor's response to the bird. Prominent critics of appraisal theories in the 20<sup>th</sup> century<sup>40</sup> argued emotion is inherently reactive and appraisals are best viewed as a consequent, certainly not a precursor, of emotional reactions.

These criticisms are largely irrelevant, however, when one views appraisal theories from a computational perspective. In our modeling work,



## Computational models of emotion are used in the design of virtual humans, autonomous embodied characters capable of face-to-face interaction with real humans through verbal and nonverbal behavior.



we address this criticism of appraisal theories by emphasizing appraisal and inferences are distinct processes that operate over the same mental representation of a person's relationship to the environment. We distinguish between the construction of this representation and its appraisal. The construction might involve slow, deliberative inferences or fast, reactive inferences. Regardless, the appraisal process that assesses this representation is fast, parallel, and automatic.

### EMA Model

As we define it, a computational model of appraisal includes: an appraisal-derivation process that interprets a representation of the person-environment relationship to derive a set of appraisal variables; an emotion-derivation model that takes this set of appraisals and produces an emotional response; and a set of behavioral consequence processes, or coping strategies, triggered by this emotion and that subsequently manipulate the person-environment. Depending on how these three processes are reified, different emotion models (corresponding to different variants of appraisal theory) can be produced; see, for example, Marsella et al.<sup>26</sup> We now turn to one approach, our own EMA (EMotion and Adaption) model<sup>17,24</sup> (see Figure 2).

**Person-environment relation.** In EMA, the model maintains an explicit representation of the “agent-environment relationship” that serves as both input to and output of the various appraisal processes. This explicit representation is the agent's view of how it relates to the environment and consists of a set of beliefs, desires, intentions, plans, utilities, and probabilities, constructs drawn from traditional notions of decision theory and AI planning. We refer to this representation as the “causal interpretation” to emphasize the importance of causal reasoning, as well as the interpretative (subjective) character of the appraisal process. The causal interpretation (which, in cognitive architecture terminology, corresponds to the agent's working memory) encodes the input, intermediate results, and output of inferential processes that mediate between an agent's goals and its physical and social environment

(such as perception, planning, explanation, and natural language processing). These inferential processes could be fast, as in recognizing the threat of a bird flying toward oneself, or more deliberative (such as forming a plan to save the bird).

The causal interpretation is a snapshot of the agent's current knowledge concerning the agent-environment relationship. This knowledge changes, moment to moment, in response to observation or inference; for example, the agent's actions or the actions of other social actors change the environment, effects that are reflected in the causal interpretation as soon as they are perceived by the agent's senses. Further, the mere act of thinking can change the perceived agent-environment relationship; for example, as the agent develops plans or forms intentions, these intentions are also reflected as changes in the causal interpretation.

The representation of the causal interpretation supports rapid, essentially reactive, assessment by the appraisal-derivation process. The need for rapid appraisal thus poses a requirement on the inferential processes that maintain the causal interpretation.

**Appraisal-derivation process.** Appraisal theories characterize emotion-eliciting events in terms of a set of specific appraisal variables, but most theories are vague with respect to the processes underlying these judgments. We assume appraisal is fast, parallel, and automatic. These characteristics are achieved by modeling appraisal as a set of continuously active feature detectors that map features of the causal interpretation into appraisal variables. All significant features in the causal interpretation are appraised separately, simultaneously, and automatically; for example, if the causal interpretation encodes an action with two consequences, one good, one bad, each consequence is appraised in parallel, and any factors influencing the desirability or likelihood of these consequences are automatically reflected in the appraisals as soon as the factors are recorded in the causal interpretation. In this sense, appraisals do not change the causal interpretation but provide a continuously updated "affective summary" of its contents; see Gratch and Marsella<sup>17</sup> for more on the architectur-



## Incorporating emotion and its expression allow virtual humans to exploit and respond to the social functions of emotion, as well as appear lifelike.



al principles distinguishing appraisal from the cognitive operations maintaining the causal interpretation.

The appraisal process in EMA associates a data structure, or "appraisal frame," with each proposition. The appraisal frame maintains a continuously updated set of appraisal values associated with each proposition. Values include:

*Perspective.* The viewpoint from which the proposition is judged; EMA appraises events from its own perspective but also from the perspectives of other agents;

*Relevance.* EMA judges propositions relevant if it has non-zero utility for some agent;

*Desirability.* The value of a proposition to the agent (such as Does it advance or inhibit its utility?) can be positive or negative;

*Likelihood.* A measure of the likelihood of propositions;

*Expectedness.* The extent to which a state could have been predicted from the causal interpretation;

*Causal attribution.* Who deserves credit/blame?;

*Controllability.* Can the outcome be altered through an agent's actions?; and

*Changeability.* Can the outcome be altered by another agent?

Each appraised event is mapped into an emotion instance of some type and intensity, as discussed next.

**Emotion derivation.** In EMA, appraisal informs an agent's coping response but is biased by an overall mood state. Mood acts as proxy for certain sub-symbolic (brain or bodily) processes<sup>40</sup> important for reconciling appraisal models with empirical observations (such as mood's influence on judgments<sup>34</sup> and core affect<sup>2</sup>). As we discussed in the bird example, such mood effects can serve an adaptive function.

At the appraisal level, EMA maintains multiple appraisal frames, one for each proposition in the causal interpretation. Multiple appraisals of even the same event are possible, as the event could affect different goals in different ways. Individual appraisal frames (and associated intensities) are aggregated into a mood, a running average of appraised events over time, disassociated from the original elicitor.

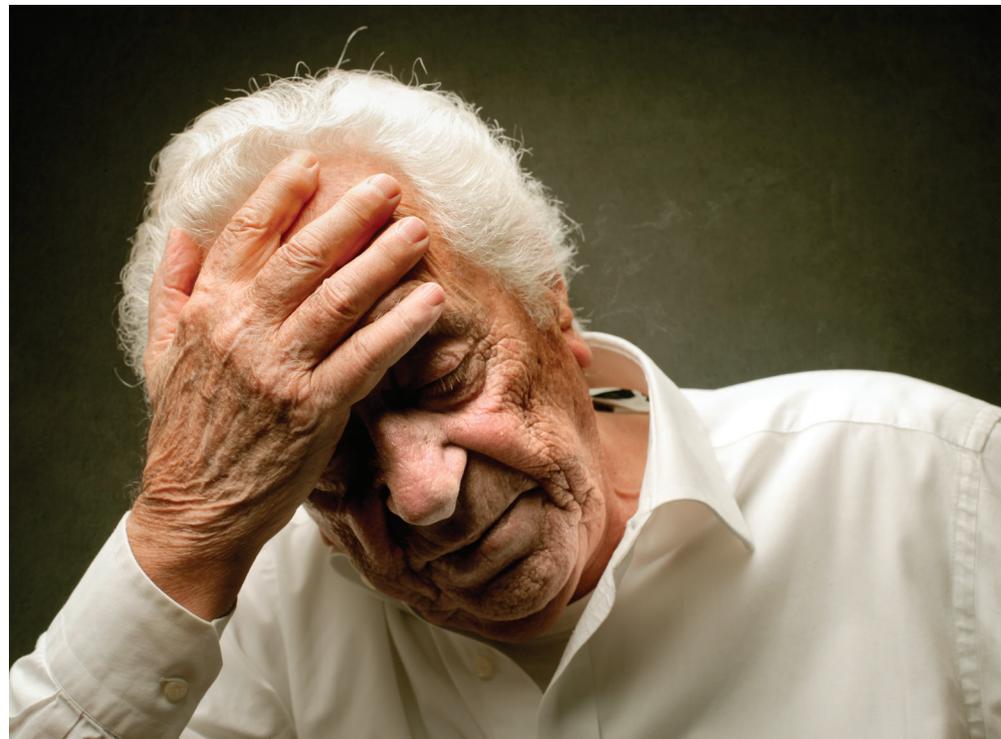
ing events; that is, mood is not intentional. EMA applies a mood adjustment to individual appraisal frames. EMA's moment-to-moment coping response is selected automatically by a simple activation-based focus of attention, with the appraisal frame that determines coping as the most recently accessed such frame with highest mood-adjusted intensity.

**Affect consequences.** Finally, EMA includes a computational model of coping in the appraisal process. Coping determines how the agent responds to the appraised significance of events, with the purpose being to alter a person's subjective interpretation of the person-environment relation. EMA proposes coping strategies to maintain desirable or overturn undesirable in-focus events, or appraisal instances. These strategies work in essentially the reverse direction of the appraisal that motivates them, by identifying features of the causal interpretation (such as beliefs, desires, intentions, and expectations) that produced the appraisal and that should be maintained or altered.

In EMA, coping strategies are control signals that enable or suppress cognitive processes that operate on the causal interpretation, including sensing, avoiding a threat, refining a plan, or adding/dropping goals and intentions. Coping sanctions actions congruent with the current appraisal pattern. EMA provides a formal realization of the coping strategies discussed in the clinical psychology literature by defining coping strategies in terms of their affect on the agent's attention, beliefs, desires, or intentions:

*Attention-related coping.* Certain coping strategies modulate the agent's attention to features of the environment, thereby altering their emotional state by altering working memory (the causal interpretation); for example, in the bird example, the actor's orientation toward the stimulus could be viewed as "seeking information" about a surprising event. Alternatively, one can "suppress information;" for example, a student concerned about a course project could avoid checking the due date.

*Belief-related coping.* A range of coping strategies suggests alterations to belief states in order to modulate these emotional states; for example, after



doing poorly on an important exam, a student might seek to "shift responsibility" to the teacher, thus alleviating self-guilt. Prior to a difficult exam, a student might cope with stress by engaging in "wishful thinking," imagining the student will do well.

*Desire-related coping.* Regulates emotion by altering goal priorities. A student facing an exam might engage in "distancing," deciding it is unimportant to do well. Or the student might "positively reinterpret" or "find a sliver lining." If passing seems impossible, well then, the student is free to party before the exam.

*Intention-related coping.* Regulates emotions by altering intentions or by taking actions; for example, our student might "plan" to study for the exam. Interestingly, the formation of an intention will alter the current emotional state in EMA, even if the plan is not executed. Alternatively, the student might cope with stress by "seeking instrumental support" from a study partner. Moreover, if the student does poorly on the exam, the student could "make amends" to his or her parents to alleviate guilt by promising to study harder next time. The student could also engage in "resignation," dropping an intention to achieve a desired state (such as believing being pre-med is desirable but unattainable).

EMA proposes strategies in parallel but adopts them sequentially, with a set of preferences resolving ties; for example, EMA prefers problem-directed strategies (such as take action, plan) if control is appraised as high, and emotion-focus strategies (such as distancing, resignation, and wishful thinking) if control and changeability are low.

Figure 1 outlines how the causal interpretation, appraisal, focus, and coping interact with one another and with the agent's perceptual and inferential processes. Recall in the bird scenario that the model's dynamics involved several sources. Events happen naturally in the world, altering the causal interpretation through perception. The agent also executes actions with effects that also alter the world and therefore causal interpretation. The agent's own inferential processes (that here are meant to include belief maintenance, action selection, and planning) can alter the causal interpretation. These changes to the causal interpretation induce changes in appraisals. Of particular importance for emotion processes, appraisal and coping operate within a loop whereby a situation may be appraised, leading to emotions and coping responses that influence the agent's inferential processes that alter the person-environment relation. This influence, in

turn, affects subsequent reappraisals of the situation. Emotional meaning thus evolves in EMA as the individual interacts with the physical and the social environment.

**Role in Theory Formation and Testing**

As noted, the relation between emotion and rational thought is a debate of long standing. One benefit of computational models is they are potentially powerful research tools that weigh on such debates by forcing explicit commitments as to how mental processes are realized, how they interrelate, and how the processes unfold over time.

EMA, in particular, makes an explicit commitment to the relation between emotion and an agent’s cognitive processes. Appraisal is treated as fundamentally reactive, an automatic assessment of the contents of mental representations. Differences in the temporal course of emotion dynamics are accordingly due to differences in the temporal course of eliciting conditions, perceptual processes, and inferential processes that maintain the representation of the person-environment relation, including both deliberative and reactive processes. This allows the model to explain in a uniform way both fast, seemingly automatic emotion responses, and slower, seemingly more deliberative responses, in contrast to more complex models and theories that postulate multiple processes.<sup>12,32</sup>

EMA assumes cognition and perception encodes the personal relevance of events in ways that make

appraisal simple, fast, and general, evolving as cognitive processes update the agent-environment relationship. Appraisal is thus not so much a process as it is an output requirement on cognition and perception. The values generated by those processes constitute the appraisal values. Similarly, we pose the inverse requirement on coping strategies, that the output of coping is the adjustment of attention, beliefs, desires, and intentions on which cognition and appraisal rely. Coping can be viewed as an inverse of the appraisal process, seeking to adjust the causal interpretation in order to alter subsequent appraisals.

These design commitments follow from our effort to address key challenges for any model of emotion. One such challenge is to explain the often rapid, seemingly reactive, dynamics of emotional process, as outlined in the bird example, that have been raised as a challenge to appraisal theories.<sup>40</sup> Another challenge is that emotions arise and evolve over a wide range of eliciting situations, from physical events to complex social situations.

In EMA, the generality of appraisal to address complex social interactions, as well as the demands of physical threats, is due largely to appraisal’s separation from a range of perceptual and cognitive processes it leverages. EMA also generalizes the role of emotion in an agent’s overall architecture. In EMA, appraisal and coping play a central role in mediating response for the agent generally. This role is in keeping with Simon’s view<sup>35</sup> of emotion as an interrupt mechanism and research<sup>4</sup>

that argues for emotion’s role in decision making.

EMA maintains that appraisal and coping shape, but do not determine, the agent’s response; for example, whether an agent copes by forming the intention to act hinges on whether its planning and problem-solving processes can identify an appropriate intention or plan. Recall the umbrella the actress was holding in the bird scenario. EMA maintains the response of preparing to whack the bird depends on the emotional response to the threat and the fact that holding the umbrella beforehand may enable (or prime) cognition’s formation of the intention.

At the same time, EMA also has significant shortcomings, thus identifying challenges for research on computational models, as well as underlying theories. Most notably, EMA’s appraisal process focuses on deriving appraisal values and treats emotion categories largely as epiphenomenal. Another limitation follows from the inferential processes that support appraisal by maintaining the causal interpretation. In particular, because the causal interpretation has limited capacity to model the beliefs of other agents, more complex social emotions (such as embarrassment) are not modeled. The inference processes also need to impose better constraints on coping strategies specifically due to the absence of constraints on belief revision; the overall model allows for wishful thinking and resignation that alters beliefs and goals while ignoring potential effects on related beliefs or goals; see Ito et al.<sup>19</sup> for a utility-based approach to address this limitation.

Nevertheless, both the strengths and weaknesses of a model like EMA support the idea of computational modeling of emotions as a powerful approach to addressing the question of the processes underlying emotion and its relation to cognition. Constructing EMA forced us to make specific commitments about the representation of the person-environment relationship, the computation of appraisals based on these representations, the role of perception, memory, interpretation, and inference in appraisal, the modeling of coping, and the relationship among appraisals, emotions, and coping. Further, once computationally re-

Figure 3. Virtual human accident scene (USC Institute for Creative Technologies).



alized, simulation allows the model to be explored systematically and manipulated, thereby generating predictions that can be validated against reactions of human subjects. Our development of the model also identified key weaknesses that must still be addressed in both model and theory.

### Role in Virtual Humans and HCI

As outlined earlier, computational models of emotion are used in the design of virtual humans, autonomous embodied characters capable of face-to-face interaction with real humans through verbal and nonverbal behavior. Incorporating emotion and its expression allow virtual humans to exploit and respond to the social functions of emotion, as well as appear lifelike.

With EMA, our goal is not simply to add emotion. Rather, we have explored how organization of a virtual human's cognitive processes around appraisal and coping can facilitate the design and integration of the multiple cognitive capabilities required to create human-like behavior, including perception, planning, dialog processing, and nonverbal communication. Appraisal theories suggest a general set of criteria and control strategies that can inform and coordinate the behavior of diverse cognitive and social functions.

Whether 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, leading to more coherent global behavior.

Consider an example from Swartout et al.<sup>37</sup> of resolving natural language ambiguities. The human participant happens on an accident scene in the virtual environment (see Figure 3); a virtual crowd has gathered; an injured boy is on the ground; a virtual soldier and troops have assembled; and there are damaged vehicles. The participant



**A model that approximates natural human behavior might be very different from one that drives a synthetic movie actor.**



asks the soldier, “What happened here?” The question is ambiguous since many things have happened: the participant just arrived; the troops assembled; an accident occurred; and a crowd formed. Whereas all these actions would in some sense be a correct response, it would be silly for the soldier to say, “You just drove up.” The expected response is a description of the accident.

To respond correctly requires determination of the linguistic focus of the discussion. A common heuristic is to use “recency,” or whatever was most recently discussed or occurred most recently. In this case, the accident scenario, recency does not work, as several things have happened subsequent to the accident.

However, people often focus most intently on what upsets them emotionally, suggesting an emotion-based heuristic for determining the focus. Because the virtual soldier incorporates EMA, the linguistic routines have access to his emotions about the accident and can use that information in determining linguistic focus, allowing the soldier to give the most appropriate answer, specifically to describe the accident and how it occurred.

Additionally, EMA models decision making as driven by appraisals of alternative plans and uses coping strategies to drive alternative responses in negotiation<sup>38</sup> (such as trying to avoid negotiation and seeking distributive solutions as opposed to integrative solutions). EMA is also used to influence the agent's beliefs; for example, under significant stress due to a blameworthy event, the virtual human can alleviate guilt or fear of reprisal by shifting its beliefs about blame.

### Validation

A computational model of emotion must be judged with respect to its intended ends. A model that approximates natural human behavior might be very different from one that drives a synthetic movie actor. A fundamental goal that drives our research is to create models that accurately predict human emotion: how it arises from the structure of situations people face; how it affects people's beliefs and actions over time; and how its manifestation affects the beliefs and behaviors of

other social actors. Here, we review several principles that guide our empirical approach to model validation; see also Staller and Petta.<sup>36</sup>

*Understanding the relationship between emotions and unfolding situations.* Although many emotion theories posit that emotions arise from individuals' relationships with their environments, most experimental work has shied away from directly manipulating the dynamic cycle of change, appraisal, and reappraisal we observed in the bird example. More common are mood-induction studies where a participant's mood is shaped independently from the situation (such as by listening to happy or sad music<sup>34</sup>) or by creating "one-shot" situations (such as watching reactions to the spinning of a roulette wheel<sup>30</sup>), as they provide a great measure of experimental control.

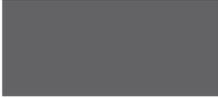
Motivated by examples of emotion-invoking scenarios (such as the bird), we have sought to develop techniques that systematically manipulate the temporal dynamics of the person-environment relationship. We have explored techniques that place laboratory participants in emotion-invoking tasks where aspects of the person-environment relationship are systematically controlled in situations that unfold and change over time, allowing us to measure moment-to-moment changes in a person's appraisals, emotions, and coping tendencies.<sup>35</sup> We then compare these responses with predictions from EMA.

*Linking intra-personal and inter-personal emotion.* As outlined in the bird scenario, emotion helps coordinate both individual and social behavior. Most empirical work adopts one perspective or the other, but we see the cognitive and social perspectives as two sources of information that mutually constrain theory and model development. We have thus sought to show the same computational models that predict the intrapersonal antecedents and consequences of emotion can also drive the social behavior of human-like agents and evoke similar social responses as those in human-to-human emotional exchanges.

One example is reverse appraisal theory,<sup>10</sup> illustrating how appraisal theory provides an explanatory



**Researchers have turned to computational models of emotion as tools to research human emotion, as well as exploit it in applications.**



framework for predicting people's responses to the emotional signals of others. It maintains that people use emotional expressions as a sort of mind reading. When observing another person's emotional reactions to an environmental event, people infer, or reverse engineer, how experiencers appraise their situation, using it to recover what goals would have led to such an appraisal. We showed empirical support for this concept by having participants play economic games with computer agents driven by the EMA model.

*Model-driven experimentation.* Finally, our work seeks to create true synergies between computational and psychological approaches to understanding emotion. We are not satisfied simply to show our models "fit" human data but rather seek to show they are generative in the sense of producing new insights or novel predictions that can inform understanding. From this perspective, computational models are simply theories, albeit more concrete ones that afford a level of hypothesis generation and experimentation difficult to achieve through traditional theories.

One example of such model-driven experimentation is seen in our work on modeling appraisals of causal attribution. One of our students created a model of the factors behind causal attribution appraisals (such as causality, intent, foreknowledge, and coercion). The model made it possible to generate hypothetically different situations that should produce different appraisals. These were then presented to human participants, and the model's predictions were consistent with human participants' responses.<sup>23</sup> Likewise, another student used appraisal theory to derive a general model of how the ways a problem is described, or framed, affects decisions people make.<sup>18</sup> The model uses appraisal variables to go beyond one-dimensional treatments of framing in terms of loss versus gain.<sup>39</sup> The model was in turn used to generate alternative decision scenarios that were presented to human subjects, and the model predictions were supported.

These principles—examining emotions in unfolding tasks, linking cognitive and social functions of

emotion, and model-driven experimentation—continue to shape our research, yet other approaches may be better suited, depending on one's underlying reason for exploring emotion; for example, models designed to inform intelligent systems might avoid some of the seemingly irrational application of coping humans adopt.

## Conclusion

Over the past half-century, there has been rapid growth of cross-disciplinary research employing computational methods to understand human behavior, as well as facilitate interaction between people and machines, with work on computational models of emotion becoming an important component. At times neglected in work in cognitive science and AI, modern research in human emotion has firmly established the powerful role emotion plays in human behavior. As a consequence, researchers have turned to computational models of emotion as tools to research human emotion, as well as exploit it in applications. Here, we have sought to illustrate the various ways such models are being used, from a communitywide general perspective to providing more specific details from our own work on EMA.

Modeling appraisal theory in an agent provides an interesting perspective on the relation between emotion and rationality. Appraisal theory argues that emotion serves to generalize stimulus response by providing more general ways to characterize types of stimuli in terms of classes of viable organism responses. For an agent, appraisal dimensions serve a general, uniform value function that establishes the personal and social significance of events. Assessments (such as desirability, coping potential, unexpectedness, and causal attribution) are clearly relevant to any social agent, whether deemed emotional or not. Having been characterized in a uniform fashion, the appraisal results coordinate systemwide coping responses that serve to guide the agent's specific responses to the eliciting event, essentially helping the agent find its ecological niche. Emotion is thus inextricably coupled to how an agent—human or artificial—reacts and responds to the world.

## Acknowledgment

The work depicted here was sponsored by the U.S. Army and the Air Force Office of Scientific Research. Statements and opinions expressed do not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred. 

## References

1. Armony, J.L., Servan-Schreiber, D., Cohen, J.D., and LeDoux, J.E. Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning. *Trends in Cognitive Science* 1, 1 (Apr. 1997), 28–34.
2. Barrett, L.F. Are emotions natural kinds? *Perspectives on Psychological Science* 1, 1 (Apr. 2006), 28–58.
3. Bechara, A., Damásio, H., Damásio, A., and Lee, G. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision making. *Journal of Neuroscience* 19, 13 (July 1, 1999), 5473–5481.
4. Becker-Asano, C. and Wachsmuth, I. Affect simulation with primary and secondary emotions. In *Proceedings of the Eighth International Conference on Intelligent Virtual Agents* (Tokyo). Springer, 2008, 15–28.
5. Blanchard, A. and Cañamero, L. Developing affect-modulated behaviors: Stability, exploration, exploitation, or imitation? In *Proceedings of the Sixth International Workshop on Epigenetic Robotics* (Paris, 2006).
6. Broekens, J. Modeling the experience of emotion. *International Journal of Synthetic Emotions* 1, 1 (Jan. 2010), 1–17.
7. Broekens, J., Kusters, W.A., and Verbeek, F.J. On affect and self-adaptation: Potential benefits of valence-controlled action-selection. In *Proceedings of the Second International Conference on Bio-inspired Modeling of Cognitive Tasks* (Manga del Mar Menor, Spain). Springer, 2007, 357–366.
8. Conati, C. and MacLaren, H. Evaluating a probabilistic model of student affect. In *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems* (Maceio, Brazil, 2004).
9. Darwin, C. *The Expression of the Emotions in Man and Animals, Third Edition*. Oxford University Press, New York, 1998.
10. de Melo, C., Gratch, J., and Carnevale, P.J. Reverse appraisal: Inferring from emotion displays who is the cooperater and the competitor in a social dilemma. In *Proceedings of the Cognitive Science Conference* (Boston, July 2011).
11. DeSteno, D., Dasgupta, N., Bartlett, M.Y., and Caidric, A. Prejudice from thin air. *Psychological Science* 15, 5 (May 2004), 319–324.
12. Dias, J. and Paiva, A. Feeling and reasoning: A computational model for emotional agents. In *Proceedings of the 12th Portuguese Conference on Artificial Intelligence* (Covilhã, Portugal). Springer, 2005, 127–140.
13. Eisenberg, N., Fabes, R.A., Schaller, M., and Miller, P.A. Sympathy and personal distress: Development, gender differences, and interrelations of indexes. In *Empathy and Related Emotional Responses New Directions in Child Development*, N. Eisenberg, Ed. Jossey-Bass, San Francisco, CA, 1989, 107–126.
14. Ekman, P. An argument for basic emotions. *Cognition and Emotion* 6, 3–4 (1992), 169–200.
15. Frank, R. *Passions With Reason: The Strategic Role of the Emotions*. W.W. Norton, New York, 1988.
16. Gebhard, P. ALMA: A Layered Model of Affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (Utrecht, the Netherlands), 2005.
17. Gratch, J. and Marsella, S. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research* 5, 4 (Dec. 2004), 269–306.
18. Ito, J.Y. and Marsella, S.C. Contextually based utility: An appraisal-based approach at modeling framing and decisions. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence* (San Francisco, Aug. 7–11). AAAI Press, 2011, 1442–1448.
19. Ito, J.Y., Pynadath, D.V., and Marsella, S.C. Modeling self-deception within a decision-theoretic framework. *Autonomous Agents and Multi-Agent Systems* 20, 1 (May 2010), 3–13.
20. Keltner, D., and Haidt, J. Social functions of emotions at four levels of analysis. *Cognition and Emotion* 13, 5

(1999), 505–521.

21. Lazarus, R.S. *Emotion and Adaptation*. Oxford University Press, New York, 1991.
22. Lester, J.C., Towns, S.G., Callaway, C.B., Voerman, J.L., and FitzGerald, P.J. Deictic and emotive communication in animated pedagogical agents. In *Embodied Conversational Agents*, J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, Eds. MIT Press, Cambridge, MA, 2000, 123–154.
23. Mao, W. and Gratch, J. Evaluating a computational model of social causality and responsibility. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (Hakodate, Japan), 2006.
24. Marsella, S. and Gratch, J. EMA: A process model of appraisal dynamics. *Journal of Cognitive Systems Research* 10, 1 (Mar. 2009), 70–90.
25. Marsella, S. and Gratch, J. Modeling the interplay of plans and emotions in multi-agent simulations. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (Edinburgh, Scotland, 2001).
26. Marsella, S., Gratch, J., and Petta, P. Computational models of emotion. In *A Blueprint for Affective Computing: A Sourcebook and Manual*, K.R. Scherer, T. Bänziger, and E. Roesch, Eds. Oxford University Press, New York, 2010, 21–46.
27. Marsella, S., Johnson, W.L., and LaBore, C. Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents* (Montréal, Canada, 2000), 301–308.
28. Mehrabian, A. and Russell, J.A. *An Approach to Environmental Psychology*. MIT Press, Cambridge, MA, 1974.
29. Panskepp, J. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, New York, 1998.
30. Reisenzein, R. Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition & Emotion* 14, 1 (2000), 1–38.
31. Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., and Meyer, J.-J.C. Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange. *IEEE Transactions on Affective Computing* 4, 3 (July 2013), 242–245.
32. Scherer, K.R. Appraisal considered as a process of multilevel sequential checking. In *Appraisal Processes in Emotion: Theory, Methods, Research*, K.R. Scherer, A. Schorr, and T. Johnstone, Eds. Oxford University Press, New York, 2001, 92–120.
33. Scheutz, M. and Sloman, A. Affect and agent control: Experiments with simple affective states. In *Proceedings of the Second Asia-Pacific Conference on Intelligent Agent Technology* (Maebashi City, Japan). World Scientific Publishing, 2001, 200–209.
34. Schwarz, N. and Clore, G.L. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45, 3 (Sept. 1983), 513–523.
35. Simon, H.A. Motivational and emotional controls of cognition. *Psychological Review* 74 (Jan. 1967), 29–39.
36. Staller, A. and Petta, P. Introducing emotions into the computational study of social norms: A first evaluation. *Journal of Artificial Societies and Social Simulation* 4, 1 (Jan. 2001), 27–60.
37. Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel, J., and Traum, D. Toward virtual humans. *AI Magazine* 27, 1 (2006).
38. Traum, D., Swartout, W., Marsella, S., and Gratch, J. Fight, flight, or negotiate. In *Proceedings of the Intelligent Virtual Agents Conference* (Kos, Greece). Springer, 2005.
39. Tversky, A. and Kahneman, D. The framing of decisions and the psychology of choice. *Science* 211, 4481 (Jan. 30, 1981), 453–458.
40. Zajonc, R.B. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35 (Feb. 1980), 151–175.

**Stacy Marsella** (stacymarsella@gmail.com) is a professor of computer science and psychology at Northeastern University, Boston, MA.

**Jonathan Gratch** (gratch@ict.usc.edu) is a professor of computer science and psychology at the University of Southern California, Los Angeles, CA.

Copyright held by owner/author(s). Publication rights licensed to ACM. \$15.00