True Emotion vs. Social Intentions in Nonverbal Communication: Towards a Synthesis for Embodied Conversational Agents

Jonathan Gratch
University of Southern California

gratch AT ict.usc.edu

Abstract. Does a facial expression convey privileged information about a person's mental state or is it a communicative act, divorced from "true" beliefs, desires and intentions? This question is often cast as a dichotomy between competing theoretical perspectives. Theorists like Ekman argue for the primacy of emotion as a determinant of nonverbal behavior: emotions "leak" and only indirectly serve social ends. In contrast, theorists such as Fridlund argue for the primacy of social ends in determining nonverbal displays. This dichotomy has worked to divide virtual character research. Whereas there have been advances in modeling emotion, this work is often seen as irrelevant to the generation of communicative behavior. In this chapter, I review current findings on the interpersonal function of emotion. I'll discuss recent developments in Social Appraisal theory as a way to bridge this dichotomy and our attempts to model these functions within the context of embodied conversational agents.

Keywords: emotion, nonverbal behavior, virtual humans, cognitive modeling

1 Introduction

Do facial expressions convey privileged information about a person's mental state or are they communicative acts, independent of a person's actual beliefs, desires and intentions? This question is often cast as a dichotomy between competing theoretical accounts of nonverbal behavior. On the one hand, theorists such as Ekman [2] or Frank [3] argue for the primacy of emotion as a determinant of nonverbal behavior: emotions "leak" through our behaviors and convey true information about our current beliefs and feelings, which only indirectly serve social ends. On the other hand, *social intentions* theorists such as Fridlund [4] or Chovil [5] argue for the primacy of social ends as the determinant of nonverbal displays: organisms use displays strategically to satisfy their social intentions and any strong connections between displays and true feelings would unnecessarily restrict their flexibility in responding to life's demands.

This dichotomy between the emotional and communicative origins of nonverbal displays is reflected in virtual character research. Whereas there has been considerable advances in modeling emotions and giving virtual characters the ability to derive situationally appropriate emotions, this work is often seen as irrelevant to the genera-

tion of communicative nonverbal behavior [6][1]. Communicative agent research has focused almost exclusively on the role of nonverbal displays in conveying propositional content or in managing conversational flow. Behaviors are typically tied to a *conversational function model* [7, 8] that represents the specific communicative functions required to support effective face-to-face conversation. Functions, such as initiating a conversational turn or emphasizing a word, are associated with nonverbal behaviors, such as looking repeatedly at another person or raising one's eyebrows, respectively. When emotional behaviors are included in such models, they are typically associated with a specific communicative function divorced from any emotional state associated with the agent. For example, Poggi and Pelachaud [9] use emotional expressions to convey the performative of a speech act, showing "potential anger" to communicate that social consequences will ensue if a request is not fulfilled. Indeed, consistent with Fridlund's view, tying emotional behaviors to an agent's motivational state (such as one exists), could limit a system designer's ability to create effective interactions (see [6]). For example, a tutoring agent that displayed frustration at a student's repeated failures could work against the goal of promoting student engagement.[2] Disassociating emotional display from motivational state could, conceivably, create superhuman agents that strategically select nonverbal displays based purely on their communicative impact, unburdened by any "emotional baggage."

The trend to discount emotion models in communicative agent research has been reinforced by the intrapsychic focus of most computational emotion research. Motivated by findings on emotion's functional role in cognitive processes, research on computational models of emotion have explored emotion's function in decision-making [12, 13], learning [14] and coping with environmental stressors [15]. These models explore the relationship between emotion and goal directed behavior, often providing detailed models of how emotions arise from a calculation of the personal significance of external events in terms of an agent's beliefs, desires and intentions. To the extent that computational models of emotion have focused on social behavior, the focus has been on general notions such as believability or empathy [16], rather than on specific socially-strategic functions. When the social function of emotion is considered at all in this work, the (typically unspoken) assumption is that model-driven behaviors can convey important social information without an explicit need to deeply model the communicative function of these behaviors (i.e., conveying true emotional states serves social ends), though this argument is rarely developed in any detail.

In this chapter, I will review current findings on the interpersonal function of emotions and their potential role in communication. I will discuss some recent developments in *social appraisal theory* [17, 18] as a way to reconcile the true emotion and

---

[1] Also Justine Cassell and Matthew Stone, personal communication.

[2] Interestingly, the closest to an integration of communicative and emotional approaches comes in the area of tutoring. Lester's COSMO system associates pedagogical communicative acts with appraisals of student performance [10]. For example, COSMO responds with admiration when the student succeeds. In my view, this approach is best seen a purely strategic model that happens to exploit a model of the student's emotions in its strategic calculations. Elliott, Rickel and Lester [11] subsequently proposed a more ambitious synthesis that begins to address the themes I explore here. Unfortunately, this work was never developed further.

social intentions views of nonverbal behavior. I will then discuss our attempts to model these functions within the context of embodied conversational agents.

2 Social Emotions

Emotions are highly social. They can arise from our understanding of the social context, impact our behavior in ways that communicate our beliefs, desires and intentions to social partners, elicit social responses that alter the social context and, thus, transform our initial emotional response. In contrast to the view that emotions interfere with rational thought, recent scholarship has emphasized the adaptive nature of many emotional responses, including their role in social interactions [19-24].

The *social functional view* of emotion emphasizes the utilitarian role of emotions in social cognition and communication [19]. This view hypothesizes that humans evolved to survive through social relationships and emotions are a fundamental building block of effective social interactions. On the one hand, emotions inform and direct cognitive processes in ways that help us successfully navigate social interactions. On the other, emotional displays influence the behaviors of others by transmitting coordinating information and eliciting adaptive social responses. In contrast to Fridlund's social intentions view, the social functional view hypothesizes a tight coupling between emotional processes and emotional behavior – we display what we feel – and this rapid, involuntary and authentic nature of emotional displays is, in fact, necessary for facilitating coordination and group cohesion.

Several findings suggest that emotions help inform and motivate social decision making. It is generally accepted that emotions help inform the individual of personally significant events and recruit the cognitive and physical resources necessary to adaptively respond. Emotions such as love, guilt, shame or anger inform us about the quality of our social relationships [19]. For example, feelings of love signal our level of commitment to another, whereas anger informs us of threats to the righteousness of our own perspective [25]. Beyond informing us of the quality of the interaction, social emotions also prepare our minds and bodies to respond to the social environment. For example, anger alters our social perceptions, sensitizing us to the injustices of others [26], and triggers physiological changes, moving blood from the internal organs towards the hands and arms in preparation of physical confrontation [19]. Finally, emotions serve as rewards or punishments, reinforcing social behavior and promoting the formation of group bonds, loyalty and identity. Trusting others actually feels good: Studies by Zak [27] suggest that acts of trust, such as cooperating on simple social games such as the Prisoner's Dilemma, lead to the release of hormones implicated in the formation of social bonds – whereas harming others feels bad – anticipatory guilt and shame help enforce social norms [20, 28].

Whereas emotion may promote adaptive social decisions, displays of emotion often promote adaptive social responses in others. On the one hand, displays provide important coordinating information to other social partners [18, 21]. Our reactions to events convey important information about our interpretations of events – an undesired stimulus might result in a frown; an unexpected one might result in an expression of surprise – and thus indirectly convey information about how we are evaluating

our situation vis-à-vis our current beliefs, desires, and intentions. They communicate our relationship with other social partners (e.g., through dominance or submission displays). On the other hand, displays of emotion seem to reflexively elicit adaptive social responses from others. Emotional behaviors are highly salient and, through affective priming [29], automatically alter perceptions and judgments. Emotional displays further trigger behavioral responses. Anger, for example, can elicit fear-related responses (even subliminal presentation) [30] or serve as a demand for someone to change the course of their interaction [31], distress can elicit sympathy [32], and joy seems to invite social interaction [33]. Many of these responses seem almost automatic and visceral and to have similar functions across a wide range of social animals [22].

Although the emotional and strategic views of emotional behavior agree that such displays help achieve social ends, they differ markedly in terms of their origins. The social functional view argues that there is considerable social utility in conveying our true feelings and that the social power of such displays, in fact, depends on their authenticity. Although there are certainly social circumstances that demand deception, arguably the vast majority of our social interactions are more mundane, where communication of true feelings and intentions is of mutual benefit. Whether you're a New Yorker navigating a busy sidewalk or a great ape finding your place in the social hierarchy, life presents us with numerous split-second decisions on how to respond to a continuously moving social landscape. In such situations, "reading" the minds of social partners can smooth interactions and avoid unnecessary "collisions." More substantial decisions, like our choice of friends, mates, or research collaborators, would yield greater mutual benefit if parties had privileged access to (at least some of) each other's beliefs and intentions. Indeed, game theorists have shown formally how so-called *commitment problems* -- where parties would receive higher utility by cooperating but choose not to based on concerns about deception (e.g., prisoner's dilemma) -- could be solved if true feelings and intentions could be divined from each other's observable behavior [3]. The social functional view of emotion argues that "authentic" emotional displays evolved because they give a selective advantage to social animals that could read each others thoughts and, thus, solve commitment problems to their mutual benefit. Consistent with this view, studies have shown that people that interact with each other before games like the prisoner's dilemma do a reasonable job at predicting the true intentions of their partners [34, 35]. This claim is further supported by findings suggesting that emotion displays play an important role in the development of mind-reading skills: discrepancies between felt emotion and other's responses to events draws attention to differences between self and other, stimulating the development of perspective taking [36], theory of mind [37] and sense of self [38]; moreover early deficits in the ability to produce or decode emotional displays may contribute to social deficits such as autism [39].

This tight, necessary, coupling between emotion and display posited by the social functional view of emotion is a major point of controversy in nonverbal behavior research and has stimulated a large body of research both supporting and contradicting this contention. Social intention theorists point to well-replicated findings illustrating how emotional displays change depending on the social context (i.e., depending on if one is alone or with friends or strangers), culture and ones social goals and intentions [40-42]. These theorists emphasize the primacy of social motives and inten-

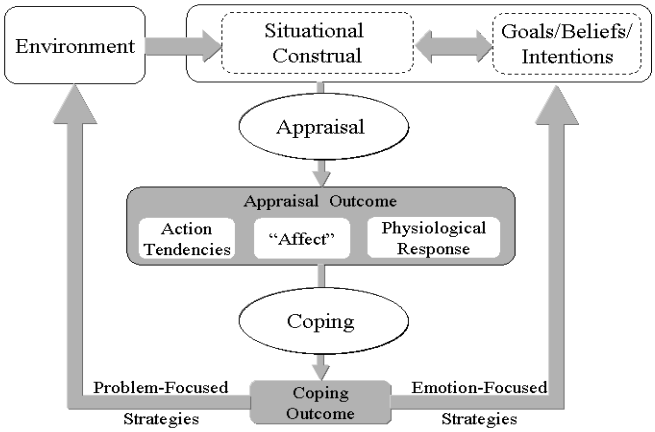

Figure 1: Schematic view of appraisal theory, adapted from Smith and Lazarus [1].

tions in determining facial expressions, downplaying or even discounting the role of emotion. Emotion theorists explain away such findings by positing a separate downstream process that filters true emotion through a set of *display rules* that attenuate the expressions depending on contextual factors [2]. In the next section, I turn to social appraisal theories of emotion that attempt to reconcile these perspectives without appeal to two-process models. Social appraisal theories emphasize the importance of the social context, social goals and intentions in our experience of emotion, suggesting this controversy arises from a false dichotomy and holding the promise of a synthesis of emotional and communicative perspectives on nonverbal behavior.

3 Social Appraisal Theory

The (false) dichotomy between communicative and emotional views of nonverbal behavior in social agents has been reinforced by the apparent intrapsychic focus of emotion theories and the computation models they have inspired. Appraisal theory [e.g., 43], a dominant theoretical perspective on emotion and view most influential in the design of cognitive models involving emotional processes, treats emotion as largely an intrapersonal, self-centered process (see [17]). Appraisal theory has heavily influenced computational research on emotion as it treats emotion as a form of information processing, providing a concrete specification of how emotion both arises from and influences cognitive processes (see Figure 1). Appraisal theory emphasizes the role of individual judgments in the elicitation and differentiation of emotional responses, specifically judgments, deemed *appraisal variables or checks*, that characterize aspects of the personal significance of 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?). Patterns of appraisal elicit emotional behav-

ior, but they also trigger stereotypical cognitive responses formalized as qualitatively distinct *coping strategies* (e.g., planning, procrastination or resignation). These responses, in turn, alter the individual's subsequent judgments concerning their relationship to the environment, emphasizing a view of emotion as a dynamic, unfolding process closely tied to cognition.

Appraisal theory is not incompatible with the social functional view of emotion – individuals can hold social goals and appraise the social consequences of events – however, many appraisal theorists, and virtually all of appraisal theory's computational derivatives, emphasize *intrapersonal* processes and the relationship between individual goals and events. For example, computational models often incorporate complex reasoning mechanisms to calculate how physical events impact an individual's beliefs or desires about physical objects, but rarely consider the inferential machinery that must underlie social judgments, such as reputation. Nor do they consider how the behavioral responses of others might influence one's evolving understanding of a social exchange, such as when an atmosphere of antagonism arises from the reciprocal exchange of postures and expressions between potential rivals [18].

Recently, some appraisal theorists have emphasized the potential of appraisal theory to encompass social goals and judgments, and thereby reconcile the apparent contradictions between the emotional and social intentions views of nonverbal displays [17, 18, 44]. At the heart of this argument is the realization that many of the factors that social intention theorists point to as mediators of nonverbal displays (e.g., social motives, social goals, power, status, and cultural norms) are crucial for assessing the personal significance of events, should influence the appraisal process, and thus should play a central role in an organism's emotional response. Thus, social appraisal theorists do not deny the importance of social intentions in communicative behavior. Rather, they emphasize the centrality of these factors in producing an emotional response, and thereby influence external behavior.

Smith et al. [17] illustrate the social appraisal perspective by considering the problems raised by social emotions such as embarrassment. Appraisal theories have tended to downplay distinctions between certain social emotions such as guilt, shame and embarrassment. However, recent scholarship has emphasized important differences regarding when displays of shame and embarrassment are elicited and the social functions they seem to serve [45]. Guilt seems to involve transgressions that violate internal standards whereas shame and embarrassment seem to arise from the perceived negative evaluations of self by others. Shame and embarrassment seem further distinguished in that shame involves situations where both self and other agree that the transgression represents a fundamental character flaw, whereas embarrassment involves a temporary condition that might be (mis)perceived as a more fundamental flaw. For example, a shy individual might find it difficult to sing in public, and feel embarrassment that their poor performance might be viewed as an inherent defect in their singing ability. Consistent with the social intentions view, these displays seem to be strategically employed. For example Leary, Landel, and Patton [46] showed that participants exhibited less embarrassment and reported feeling less embarrassed after singing aloud in public when they believed that the experimenter already knew they were uncomfortable. Leary et al. [46] argue this result illustrates that embarrassment serves self-presentational ends and therefore ceases to be necessary once these ends are fulfilled.

Adopting a social appraisal perspective, Smith et al. [17] argue that distinctions between emotions such as shame and embarrassment can be handled within appraisal theory with some greater attention to social goals and social inferences. Embarrassment involves the appraisal of a social goal –i.e., reputation or self-presentation – which may or may not be threatened, depending on the inferences formed by others. Although requiring more complex inference than typically considered in computational models of appraisal, such social appraisals could be incorporated into computational appraisal models, facilitating a synthesis of communicative and emotional views of nonverbal behavior. I now consider the implications of this perspective within our own work on the EMA computational appraisal model.

4 Computational Models

EMA is a computational model of the cognitive antecedents and consequences of emotion as posited by appraisal theory [15, 47, 48]. As in many computational models based on appraisal theory, EMA emphasizes the role of individual goals and judgments in emotional processes.[3] Here I discuss the extent to which EMA supports social appraisal theory and suggest some straightforward extensions that can increase the relevance of such models to designers of communicative agents.

In translating appraisal theory into a concrete computational model of emotional processes, EMA draws extensively on common artificial intelligent methods of reasoning and representation. EMA must represent the agent's relationship to its environment and capture the dynamics of processes involved in interpretation (appraisal) and manipulation (coping) of this representation. To this end, EMA represents the relationship between events and an agent's internal beliefs desires and intentions by building on AI planning to represent the physical relationship between events and their consequences, and BDI frameworks to represent the epistemic factors that underlie human (particularly social) activities.

Appraisal processes characterize this representation in terms of individual appraisal judgments. These extend traditional AI concerns with utility and probability:

- Desirability: what is the utility (positive or negative) of the event if it comes to pass.
- Likelihood: how probable is the outcome of the event.
- Causal attribution: who deserves credit/blame.
- Controllability: can the outcome be altered by actions under control of the agent.
- Changeability: can the outcome change on its own.

Patterns of appraisal elicit emotional displays, but they also initiate coping processes to regulate the agent's cognitive response to the appraised emotion. Coping strategies work in the reverse direction of appraisal, identifying plans, beliefs, desires or intentions to maintain or alter. These include "problem focused" strategies (e.g. planning) directed towards improving the world (the traditional concern of AI techniques) but

[3] EMA is an evolution of earlier computational models that anticipated some of the social appraisals discussed below (see [see 49]).

also encompasses "emotion-focused" strategies that impact an agent's epistemic and motivational state:

- 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
- Denial: lower the perceived likelihood of an undesirable outcome
- Mental disengagement: lower utility of desired state
- Shift blame: shift responsibility for an action toward some other agent

Strategies give input to the cognitive processes that actually execute these directives. For example, planful coping generates an intention to act, leading 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.

EMA uses an explicit representation of plans, beliefs, desires and intentions to capture output and intermediate results of processes that relate the agent to its physical and social environment. This represents the agent's current view of the agent-environment relationship, which changes with further observation or inference. EMA treats appraisal as a mapping from syntactic features of this representation to individual appraisal variables. Multiple appraisals are aggregated into an overall emotional state that influences behavior. Coping directs control signals to auxiliary reasoning modules (i.e., planning, or belief updates) to overturn or maintain features of the representation that lead to individual appraisals. For example, coping may abandon a cherished desire in response to an uncontrollable threat.

Even with its intrapsychic focus, EMA has a broad effect on behavior of the embodied conversational agents with which it has been integrated [50, 51]. Mentally, it impacts their communicative motives and biases their interpretation of ambiguous events, including the user's speech. Physically, it affects what agents say and the accompanying gestures, postures and facial expressions.

4.1 Social Appraisals

Certain appraisal checks identified by appraisal theory presuppose some form of social inference. Specifically, emotions such as anger involve an appraisal of *causal* attribution – i.e., which causal agent is responsible and blame/creditworthy for an action. Appraisal theories differ in the level of detail to which they consider causal attributions but researchers that have examined such appraisals closely emphasize that they involve complex social inferences and theory of mind. People rarely use simple causal interpretations when explaining social actions. In contrast to how causality is used in the physical sciences, the judgment of responsibility is a multi-step process involving judgments of causality, foreseeability, intention, coercion and excuse [52, 53]. As in other appraisal checks, social attributions involve evaluating consequences of events with personal significance to an agent. The evaluation is always from a perceiving agent's perspective and the significance of the consequences is based on an individual perceiver's preferences. The perceiver uses her own knowledge about

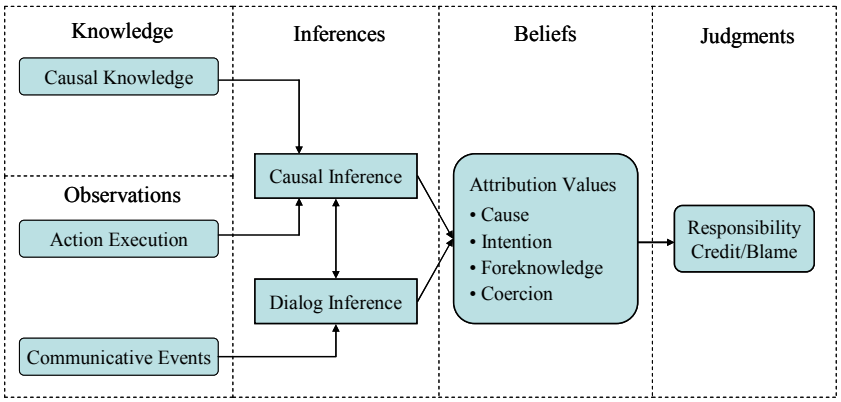

Figure 2: Computational model of causal attribution

the observed agents and her observations to form beliefs about others knowledge, beliefs and motives when forming an overall judgment.

Although most computational appraisal models use simple criteria to assess causal attribution – and thereby sidestep the need for complex social judgments – Mao and Gratch have developed and empirically validated a detailed model of how people form such judgments (Figure 2), including not only causal factors, but also epistemic variables such as freedom of choice, intention and foreknowledge [54, 55]. As a result, an actor may physically cause an event, but be absolved of responsibility and blame, or conversely, blamed for what she did not physically cause.

Building on the causal representation used by EMA, which provides a concise description of the causal relationship between events and states and a clear structure for exploring alternative courses of actions, the model forms key inferences necessary to form attributions: recognizing the relevance of events to an agent's goals and plans (key for intention recognition), assessing an agent's freedom and choice in acting (key for assessing coercive situations) and detecting how an agent's plan facilitates or prevents the plan execution of other agents (key for detecting plan interventions).

The model derives attribution variables via inference over prior causal knowledge and observations, and uses beliefs of the variables to form an overall judgment. Two important sources of information contribute to the inference process. One source is the actions performed by the agents in social events (including physical acts and communicative acts). The other is the general causal knowledge about actions and world states. Causal inference reasons about beliefs from causal evidence. Dialogue inference derives beliefs from communicative evidence. Both inferences make use of commonsense knowledge and generate beliefs of attribution variables. These beliefs serve as inputs for the attribution process, which is described as an algorithm in our computational model. Finally, the algorithm forms an overall judgment and assigns proper credit or blame to the responsible agents.

4.2 Appraising Speech Acts and Communicative Goals

One limitation of EMA is its emphasis on physical actions and physical causality. EMA reasons about the relationship between physical actions and physical goals, allowing agents to appraise the individual and social consequence of past or future actions. For example, imagine that an EMA agent is about to defect against its cooperating partner in the Prisoner's Dilemma game. The model can reason that this action will achieve the agent's goal of maximizing monetary reward (leading to positive emotions) but harm the partner (leading the partner to experience negative emotions) and possibly involve negative personal emotions (such as guilt). User speech only impacts appraisal indirectly, as when a communication leads changes an agent's representation of physical states or actions.

Linguists have long argued that speech can be viewed as a form of action with consequences that can achieve or threaten goals [56]. Although nothing prevents one from encoding speech acts and communicative goals within the causal representations used by EMA (e.g., the SAY_I_DEFECT action has the effects I_WIN and YOU_LOSE), in practice, we have maintained separate representations for physical and communicative acts and goals in our applications [50, 51]. As a result, the agent doesn't directly understand the emotional consequences of its speech acts. Rather, an external dialogue manager produces the speech act and records its causal and epistemic consequences. Only then are these consequences made available to appraisal. In this sense, the agent doesn't anticipate and mitigate the emotional consequences of its speech, as posited, for example, by politeness theory [57].

Rather, appraisal models like EMA could use a uniform representation for physical and communicative acts. This will allow, for example, the agent to anticipate the social consequences of its speech and, for example, adopt mitigation strategies to deflect any negative effects for social partners (e.g., [58]).

4.3 Social Goals

EMA, like many appraisal models, reasons about the relationship between events and goals but doesn't make explicit distinctions between classes of goals. On the other hand, social appraisal theory argues that there are different classes of goals depending on the social considerations they involve. Some goals are purely individual and don't depend on other social actors for their satisfaction. Other goals are inherently social. For example, imagine an agent choosing whether to defect in a prisoner's dilemma game. This choice presumably impacts its individual goal of gaining material resources. It might threaten a social goal of acting fairly towards others. Finally, its choice might impact "second-order" goals such as "I want my partner to believe I'm fair." Social appraisal theories argue that threats to these goals lead to very different emotions: guilt or shame, respectively [17].

One place the connection between social goals and emotion has been explicitly considered is within the intelligent tutoring literature. For example, Elliott, Rickel and Lester [11] identify several goals a tutoring system should have including giving effective explanations about the subject domain, engaging the student, ensuring the student retains task knowledge, ensuring student exhibits proper caution towards

hazardous situations. These goals are inherently social in that they cannot be achieved by the agent alone but requires cooperation by the student as well. Elliott et al propose specific emotional signals that might motivate the student to help achieve these goals

In expanding such work, one organizing concept could be to map out how goals differ with respect to the social manipulations required to establish or repair them. For example, individual goals can be achieved by purely individual acts. Joint goals (goals where both parties must contribute to its achievement) may require actions that impact the other party's motivational state. Second order goals (e.g., I want you to think I'm smart), can ignore the other party's motivations but must influence their belief state. Social appraisal theory argues that emotional displays of guilt, shame and embarrassment achieve these targeted social manipulations – though whether a communicative agent can implement and achieve these targeted effects is a fascination empirical question, as yet untested.

4.4 Social Perception and Reactivity

Most computational models of emotion de-emphasize the interactive nature of social emotions. Whether they are derived from an underlying appraisal model or arise from a communicative goal, computational systems typically follow a simple "transmission model" [59] in that a discrete emotion is calculated and transmitted faithfully to a human observer. In contrast, many contemporary theories emphasize that emotions can be dynamic and partial, reflecting the organism's transitory understanding of the physical and social context which continuously evolves from changes in the environment, inference, and social feedback. For example, feelings of antagonism may arise as facial expressions and postures are exchanged, as "one person's gradual leaning forward first leads to withdrawal until ground is held" [18]. In such a fashion, feelings may arise, not from each individual's understanding of how the interaction relates to their personal beliefs and desires, but is "co-constructed" from moment-to-moment feedback with a conversational partner [60]. This can be seen as analogous to arguments over deliberate vs. situated models of activity [61, 62]: as Herb Simon observed that the complexity of an ant's movements arise from the interaction of simple goal-directed behavior with a complex environment [63], complex emotional responses can arise from the interaction of social goals and the dynamic social environment.

Before an agent can "co-construct" emotional experiences with a human, it must perceive and respond to moment-to-moment changes in the social and physical environment. Unfortunately, most "embodied" conversational agents have a peculiar and unidirectional notion of embodiment. Although they have graphical eyes, they cannot see: most agents cannot detect the gaze, gestures, facial expressions or posture of their conversation partner. Though they have virtual ears, they cannot hear: most agents cannot detect the prosody or emotion in a human's voice. Rather, an agent receives a transcription of a speaker's words several hundred milliseconds after it was uttered. Thus, they are incapable of providing the rapid moment-to-moment emotional feedback seen in natural conversations [64]. When conversational agents do extract propositional and interactional (e.g., turn taking) meaning from the user's speech (after unnatural delay), rarely do they attempt to calculate the type of social appraisals discussed above, and which humans seem to perform effortlessly.

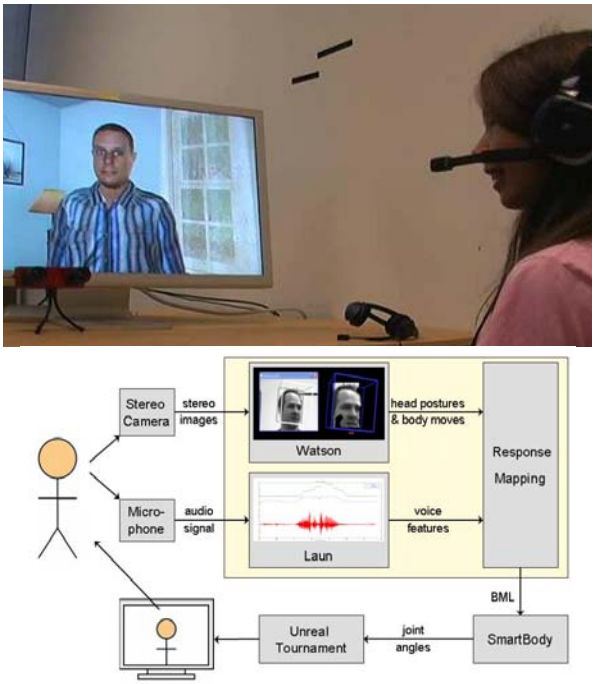

Figure 3: A child telling a story to the RAPPORT AGENT

These limitations are beginning to change. Several technologies now support the real-time analysis of prosody [65], facial expressions [66], gaze and gestures [67]. In a few instances, these technologies have been integrated into complete agent systems that can rapidly respond to a user's nonverbal behaviors [68-71], although the integration is shallow when compared to the capabilities of most embodied conversational systems. For example, the RAPPORT AGENT [68] can appear to listen to user's monolog by providing properly timed listening feedback triggered by the speaker's gestures and prosody. This has a demonstrable beneficial effect on speaker engagement and speech fluency even though the RAPPORT AGENT does not recognize or react to the content of user speech. For example, it uses the algorithm of Ward and Tsukahara (2000) to nod in response to certain prosodic contours. Such "tricks" can work in suitably constrained contexts but only scratch the surface of how to utilize multi-modal recognition technology.

I see two immediate ways to exploit perceptual information to improve the socio-emotional capabilities of agents. A "top-down" approach would be to incorporate incremental perceptual evidence into an agent's mechanisms for appraising the social context. Following the assumptions of the social functional view of emotion, such nonverbal signals should be diagnostic of the person's evaluation of events, and attending to them should improve the agent's understanding of the user's mental state. This could facilitate, for example, a sort of socio-emotional hypothesis testing: whenever an agent performs an act (physical or speech) that is relevant to another, the

agent predicts how the other would appraise the act, then uses the observed response to confirm/disconfirm the agent's theory of the other's mind. More specifically, an agent might wish to provide helpful advice to a student. If the student scowls in response, this may indicate the agent's user model is incorrect or that it has failed to follow some social norm concerning politeness and autonomy [72].

A complementary approach would be to work "bottom-up" to extend the sophistication of the more reactive sensory-motor strategies such as the RAPPORT AGENT. Jondottir et al. [64] describe a preliminary attempt along these lines. They incorporated continuous speech recognition to provide emotional feedback when the user describes certain emotionally-salient events in sexual harassment video. The system was able to recognize these events with some accuracy, but couldn't come close to responding with the rapidity of human listeners. For example, when speakers describe a scene involving an inappropriate demand for a foot massage, most human listeners responded very rapidly with facial expressions indicating surprise (within 350msec on average). Although the system was able to recognize the term "foot massage," response times were on the order of 1 to 3 seconds, a highly noticeable delay. Nevertheless, it should be possible to extend the range of situations where a system can approximate "deep" feedback with shallow features by detecting shallow semantic and contextual features.

Though potentially powerful, such strategies must recognize the inherent complexity of emotional behavior. Whether nonverbal displays arise from social motives or true emotion, they depend on a myriad of unobservable factors including the individual's goals, beliefs, and understanding of the social context. Ultimately, systems that can integrate both cognitive and sensory-motor strategies will be needed to co-construct interactive emotional experiences.

5 Conclusion

I have argued that the conflicting views on the origins of nonverbal displays – that emotional displays either reflect true emotion or are tied to achieving specific social goals divorced from emotion – arise from a false dichotomy. Rather, both perspectives are largely correct: emotional displays convey social motives and social intentions, but such motives and intentions are key constituents of true emotion. Emotions help inform the organism how it is doing in the world and recruit the cognitive and physiological resources required to adaptively respond to threats and opportunities. As social animals, it should not be surprising that these threats and opportunities arise from our social needs and demand some form of communication and coordination with other social agents. Computational models of emotion have made considerable progress in allowing asocial agents to calculate the personal significance of events and adaptively respond. By suggesting how to extend these methods toward more social inferences, I hope to have, in some measure, bridged the gap between emotional and communicative methods to generate nonverbal behavior in embodied conversational agents.

Acknowledgements. The ideas expressed here grow out of discussions on the nature of nonverbal displays with Nicole Krämer and reflections on joint work on social attributions with Wenji Mao, and would not be possible without those collaborations. Wendy Treynor reviewed a version of the draft. 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

1. Smith, C.A. and R. Lazarus, Emotion and Adaptation, in *Handbook of Personality: theory & research*, L.A. Pervin, Editor. 1990, Guilford Press: NY. p. 609-637.
2. Ekman, P., Universals and cultural differences in facial expressions of emotions, in *Nebraska Symposium on Motivation*, J. Cole, Editor. 1972, University of Nebraska Press: Lincoln, Nebraska. p. 207-283.
3. Frank, R.H., Introducing Moral Emotions into Models of Rational Choice, in *Feelings and Emotions*, A. Manstead, N. Frijda, and A. Fischer, Editors. 2004, Cambridge University Press: Cambridge. p. 422-440.
4. Fridlund, A., *Human facial expression: An evolutionary view*. 1994, San Diego, CA: Academic Press.
5. Chovil, N., Social determinants of facial displays. Journal of Nonverbal Behavior, 1991. **15**: p. 163-167.
6. Krämer, N.C., I. Iurgel, and G. Bente. Emotion and motivation in embodied conversational agents. in Proceedings of the Symposium "Agents that Want and Like", Artificial Intelligence and the Simulation of Behavior (AISB). 2005. Hatfield: SSAISB.
7. Cassell, J., et al., Human conversation as a system framework: Designing embodied conversational agents, in *Embodied Conversational Agents*, J. Cassell, et al., Editors. 2000, MIT Press: Boston. p. 29-63.
8. Heylen, D. Challenges Ahead. Head Movements and other social acts in conversation. in AISB. 2005. Hertfordshire, UK.
9. Poggi, I. and C. Pelachaud, Emotional Meaning and Expression in Performative Faces, in *Affective Interactions: Towards a New Generation of Computer Interfaces*, A. Paiva, Editor. 2000, Springer-Verlag: Berlin. p. 182-195.
10. Lester, J.C., S.G. Towns, and P.J. FitzGerald, Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. International Journal of Artificial Intelligence in Education, 1999. **10**(3-4): p. 278-291.
11. Elliott, C., J. Rickel, and J. Lester, Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis, in *Artificial Intelligence Today: Recent Trends and Developments*, M. Wooldridge and M. Veloso, Editors. 1999, Springer: Berlin/Heidelberg. p. 195-212.

12. Lisetti, C. and P. Gmytrasiewicz, Can a rational agent afford to be affectless? A formal approach. Applied Artificial Intelligence, 2002. **16**: p. 577-609.
13. Scheutz, M. and A. Sloman. Affect and agent control: experiments with simple affective states. in IAT. 2001: World Scientific Publisher.
14. Blanchard, A. and L. Cañamero. Developing Affect-modulated behaviors: stability, exploration, exploitation, or imitation? in 6th International Workshop on Epigenetic Robotics. 2006. Paris.
15. Marsella, S. and J. Gratch. Modeling coping behaviors in virtual humans: Don't worry, be happy. in Second International Joint Conference on Autonomous Agents and Multi-agent Systems. 2003. Melbourne, Australia.
16. Neal Reilly, W.S., Believable Social and Emotional Agents. 1996, Carnegie Mellon University: Pittsburgh, PA.
17. Smith, C.A., B. David, and L.D. Kirby, Emotion-Eliciting Appraisals of Social Situations, in *Affective Influences on Social Behavior*, J. Forgas, Editor. 2006, Psychology Press: New York.
18. Parkinson, B., Putting appraisal in context, in *Appraisal processes in emotion: Theory, methods, research*, K. Scherer, A. Schorr, and T. Johnstone, Editors. 2001, Oxford University Press: London. p. 173-186.
19. Keltner, D. and J. Haidt, Social Functions of Emotions at Four Levels of Anysis. Cognition and Emotion, 1999. **13**(5): p. 505-521.
20. Barrett, K.C., A functionalist approach to shame and guilt, in *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*, J.P. Tangney and K.W. Fischer, Editors. 1995, Guilford Publications: New York:. p. 25-63.
21. Spoor, J.R. and J.R. Kelly, The evolutionary significance of affect in groups: communication and group bonding. Group Processes and Intergroup Prelations, 2004. **7**(4): p. 398-412.
22. de Waal, F., *Primates and Philosophers: How Morality Evolved*, ed. S. Macedo and J. Ober. 2006, Princeton, NJ: Princeton University Press.
23. Frank, R., *Passions with reason: the strategic role of the emotions*. 1988, New York, NY: W. W. Norton.
24. Lutz, C. and G.M. White, The anthropology of emotions. Annual Review of Anthropology, 1986. **15**: p. 405-436.
25. Clore, G.L. and J. Storbeck, Affect as Information about Liking, Efficacy, and Importance, in *Affective Influences on Social Behavior*, J. Forgas, Editor. 2006, Psychology Press: New York.
26. Tiedens, L.Z., The Effect of Anger on the Hostile Inferences of Aggressive and Nonaggressive People: Specific Emotions, Cognitive Processing, and Chronic Accessibility Motivation and Emotion, 2001. **25**(3).
27. Zak, P.J., Neuroeconomics. Philosophical Transactions of the Royal Society Biology, 2004. **359**: p. 1737-1748.
28. Izard, C., *Human emotion*. 1977, New York: Plenum.
29. Klauer, K.C. and J. Musch, Affective priming: Findings and theories, in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, K.C. Klauer, J. Musch, and N.J. Mahwah, Editors. 2003, Lawrence Erlbaum. p. 7-49.

30. Dimberg, U. and A. Öhman, Beyond the wrath: Psychophysiological responses to facial stimuli. Motivation and Emotion, 1996. **20**: p. 149-182.
31. Emde, R.N., T.J. Gaensbauer, and R.J. Harmon, Emotional expression in infancy: A biobehavioral study. Psychological Issues, 1976. **10**(1, Monograph No. 37).
32. Eisenberg, N., et al., Sympathy and personal distress: Development, gender differences, and interrelations of indexes, in *Empathy and related emotional responses* N. Eisenberg, Editor. 1989, Jossey-Bass: San Franciso. p. 107-126.
33. Haviland, J.M. and M. Lelwica, The induced affect response: 10-week-old infants' responses to three emotion expressions. Developmental Psychology, 1987. **23**: p. 97-104.
34. Brosig, J., Identifying cooperative behavior: some experimental results in a prisoner's dilemma game. Journal of Economic Behavior and Organization, 2002. **47**: p. 275-290.
35. Frank, R.H., T. Gilovich, and D.T. Regan, The evolution of one-shot cooperation: an experiment. Ethology and Sociobiology, 1993. **14**: p. 247-256.
36. Dunn, J., Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. Cognition and Emotion, 1995. **9**: p. 187-201.
37. Harris, P.L., *Children and Emotion*. 1989, Oxford: Blackwell.
38. Eder, R.A., Uncovering young children's psychological selves: Individual and developmenal differences. Child Development, 1990. **61**: p. 849-863.
39. Phillips, M.L., Facial processing deficits and social dysfunction: how are they related? Brain, 2004. **127**(8): p. 1691-1692.
40. Parkinson, B., Do facial movements express emotion or communicate motives? Personality and Social Psychology Review, 2005. **9**(4): p. 278-311.
41. Manstead, A., A.H. Fischer, and E.B. Jakobs, The Social and Emotional Functions of Facial Displays, in *The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction)*, P. Philippot, R.S. Feldman, and E.J. Coats, Editors. 1999, Cambridge Univ Press. p. 287-316.
42. Fernández Dols, J.M. and M.-A. Ruiz-Belda, Spontaneous facial behavior during intense emotional episodes: artistic truth and optical truth, in *The Psychology of Facial Expression*, J.A. Russell and J.M. Fernández Dols, Editors. 1997.
43. Scherer, K.R., A. Schorr, and T. Johnstone, eds. Appraisal Processes in Emotion. Affective Science, ed. R.J. Davidson, P. Ekman, and K.R. Scherer. 2001, Oxford University Press.
44. Parkinson, B., A.H. Fischer, and A.S.R. Manstead, *Emotion in social relations: Cultural, group, and interpersonal processes*. 2005, New York: Psychology Press.
45. Tangney, J.P., Assessing individual differences in proneness to shame and guilt: development of the self-conscious affect and attribution inventory. Journal of Personality and Social Psychology, 1990. **59**: p. 102-11.
46. Leary, M.R., J.L. Landel, and K.M. Patton, The motivated expression of embarrassment following a self-presentational predicament. Journal of Personality and Social Psychology, 1996. **64**: p. 619-636.

47. Gratch, J. and S. Marsella. Tears and Fears: Modeling Emotions and Emotional Behaviors in Synthetic Agents. in Fifth International Conference on Autonomous Agents. 2001. Montreal, Canada: ACM Press.
48. Gratch, J. and S. Marsella, A domain independent framework for modeling emotion. Journal of Cognitive Systems Research, 2004. **5**(4): p. 269-306.
49. Marsella, S., J. Gratch, and J. Rickel, Expressive Behaviors for Virtual Worlds, in *Life-like Characters Tools, Affective Functions and Applications*, H. Prendinger and M. Ishizuka, Editors. 2003, Springer-Verlag: Berlin. p. 317-360.
50. Swartout, W., et al., Toward Virtual Humans. AI Magazine, 2006. **27**(1).
51. Rickel, J., et al., Toward a New Generation of Virtual Humans for Interactive Experiences, in *IEEE Intelligent Systems*. 2002. p. 32-38.
52. Weiner, B., Responsibility for Social Transgressions: An Attributional Analysis, in *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin, Editors. 2001, The MIT Press.
53. Shaver, K.G., *The attribution of blame: Causality, responsibility, and blameworthiness*. 1985, NY: Springer-Verlag.
54. Mao, W. and J. Gratch. Evaluating a computational model of social causality and responsibility. in 5th International Joint Conference on Autonomous Agents and Multiagent Systems. 2006. Hakodate, Japan.
55. Mao, W. and J. Gratch. Social Judgment in Multiagent Interactions. in Third International Joint Conference on Autonomous Agents and Multiagent Systems. 2004. New York, New York.
56. Austin, J., *How to Do Things with Words*. 1962: Harvard University Press.
57. Brown, P. and S.C. Levenson, *Politeness: Some Universals in Language Usage*. 1987, New York, NY: Cambridge University Press.
58. Martinovski, B., et al. Mitigation Theory: An integrated approach. in Cognitive Science. 2005.
59. Shannon, C.E. and W. Weaver, *A mathematical theory of communication*. 1948, Urbana, IL: University of Illinois Press.
60. Fogel, A., *Developing through relationships: Origins of communication, self and culture*. 1993, New York: Harvester Wheatsheaf.
61. Agre, P. and D. Chapman. Pengi: an implementation of a theory of activity. in National Conference on Artificial Intelligence. 1987.
62. Suchman, L.A., *Plans and situated actions: The problem of human-machine communication*. 1987, New York: Cambridge University Press.
63. Simon, H., *The sciences of the Artificial*. 1969, Cam bridge, MA: MIT Press.
64. Jondottir, G.R., et al. Fluid Semantic Back-Channel Feedback in Dialogue: Challenges and Progress. in Intelligent Virtual Agents. 2007. Paris, France: Springer.
65. Ward, N. and W. Tsukahara, Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics, 2000. **23**: p. 1177-1207.
66. Bartlett, M.S., et al. Fully automatic facial action recognition in spontaneous behavior. in 7th International Conference on Automatic Face and Gesture Recognition. 2006. Southampton, UK.

67. Morency, L.-P., et al. Contextual Recognition of Head Gestures. in 7th International Conference on Multimodal Interactions. 2005. Torento, Italy.
68. Gratch, J., et al. Virtual Rapport. in 6th International Conference on Intelligent Virtual Agents. 2006. Marina del Rey, CA: Springer.
69. Breazeal, C. and L. Aryananda, Recognition of Affective Communicative Intent in Robot-Directed Speech. Autonomous Robots, 2002. **12**: p. 83-104.
70. Gratch, J., et al., Creating Rapport with Virtual Agents, in *7th International Conference on Intelligent Virtual Agents*. 2007: Paris, France.
71. Burleson, W., Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance. 2006, Unpublished PhD Thesis, MIT Media Lab: Boston.
72. Wang, N., et al. Experimental evaluation of polite interaction tactics for pedagogical agents. in Intelligent User Interfaces. 2005.