Combining Cognitive Appraisal and Sensors for Affect Detection in a Framework for Modeling User Affect

Cristina Conati

Introduction

Digital games are one of the most promising media for the development of innovative educational content (e.g., de Castell & Jenson, 2007; Gee, 2003). They integrate game design concepts with instructional design techniques in order to better address the learning needs of this generation, which highly regards interactive, experiential learning. While there is ample evidence that educational games (edu-games from now on) are more appealing than traditional learning environments, there is still limited empirical research that supports evidentiary claims about what is learned through play, what are the pedagogical and new media constructs required to have games that teach, and what is the interplay between entertainment and learning.

In our research, we have addressed this problem by hypothesizing that a key role in edu-game effectiveness is played by learners' individual differences, both long-term (e.g., preexisting knowledge, personality traits) and short-term (e.g., interaction goals, emotional state, learning state). The more the edu-game understands about its current learner, the better it can *adapt* the interaction to fit the learner's needs. In particular, by monitoring both the learner's affective states and his or her learning trajectory, a *user-adaptive* edu-game should be better able to strike the right balance between instruction and entertainment, leveraging the latter when there is need to revive the learner's motivation and engagement.

In the context of this research, we have been investigating how to build user models that can help an educational game understand how to best support a profitable interaction with the learners. We then use these models to experiment on how to provide user-adaptive interventions that address the interplay of affect and learning.

C. Conati (✉)
Department of Computer Science, University of British Columbia,
2366 Main Hall, Vancouver, BC V6T 1Z4, Canada
e-mail: conati@cs.ubc.ca

R.A. Calvo and S.K. D'Mello (eds.), *New Perspectives on Affect and Learning Technologies*,
Explorations in the Learning Sciences, Instructional Systems and Performance Technologies 3,
DOI 10.1007/978-1-4419-9625-1_6,

A distinguishing feature of our research is that we are looking at affective user models that rely on an explicit representation of both the *potential causes* of a user affective reaction, as well as the *behavioral effects* of that reaction. The advantage of such a model is twofold. First, by relying on both causes and effects as sources of information, the model is more resilient to limitations of each individual source and thus it can more accurately assess *which* affective state the user is in. Second, by having an explicit representation of *why* the student is in a given emotional state, the model provides the game with valuable additional information to decide how to react to that state, if necessary.

Most existing efforts to recognize user affect have either relied solely on detection of behavioral reactions (e.g., Healey & Picard, 2005; Prendinger, Mori, & Ishizuka, 2005), or have blended context and effect information as features to build classifiers that predict the user emotion but cannot tell why they occur (e.g., Cooper, Muldner, Arroyo, Park Woolf, & Burleson, 2010; D'Mello & Graesser, 2010).

In the rest of this chapter, we first describe the general approach and its theoretical underpinnings. Next, we describe the Prime Climb game, the testbed we have been using to apply the framework in practice. We then introduce the general steps needed to build an affective user model following our approach, and discuss how these steps were implemented to build the affective model for Prime Climb. We conclude by reporting results on the model's performance, followed by a discussion of future work.

The Affect-Modeling Framework

Our approach relies on Dynamic Decision Networks (DDN) to leverage information on both the possible causes and the observable effects of the user's affective reaction. Figure 1 shows a high-level representation of two time-slices in our DDN-based framework for modeling user affect (Conati, 2002). Each time-slice represents the system belief over relevant elements of the world after an interaction event of interest, such as a user's action (left slice) or an action from an interface agent (right slice). As the figure shows, the network can combine evidence on both the causes and effects of emotional reactions to assess the user's emotional state after each event. Links between variables in different time-slices represent relevant temporal dependencies, such as permanence or decay.

The subnetwork above the nodes *Emotional States* is the predictive component of the framework, representing the relations between emotional states and their possible causes as described in the OCC cognitive theory of emotions (Ortony, Clore, & Collins, 1988). According to this theory, emotions derive from one's appraisal of the current situation (consisting of events, agents, and objects) with respect to one's goals, preferences, and attitudes. For instance, depending on whether an event (e.g., the outcome of an interface agent's action) fits or does not fit with one's goals, one will feel either joy or distress in relation to the event. If the current event is caused by a third-party agent, one will feel admiration or reproach

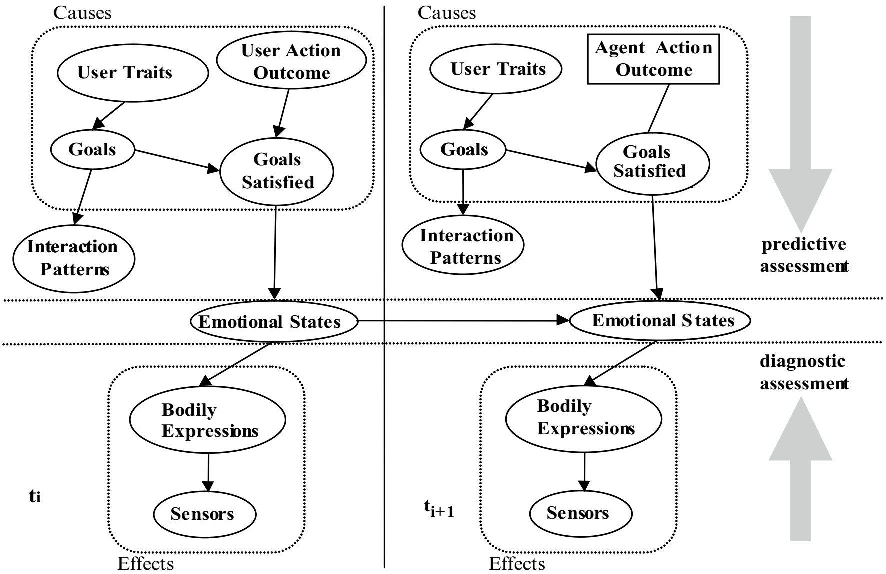

Fig. 1 High-level representation of the Dynamic Decision Networks (DDN) for affective user modeling

toward the agent; if that agent is oneself, one will feel either pride or shame. Based on this structure, the OCC theory defines 22 different emotions, which are inherently linked to context, and modulated both by factors more cognitive in nature (e.g., goals) as well as by affective elements such as attitudes and dispositions (e.g., liking/disliking not necessarily justified by objective reasons).

We based our model on the OCC theory because its intuitive representation of the causal nature of emotions lends itself well to devising computational models that can assess not only which emotions a user feels, but also why. Our OCC-based DDN includes variables for goals that a user may have during the interaction with a system that includes an interface agent (nodes *Goals* in Fig. 1). The events subject to the user's appraisal are the outcomes of the user's or the agent's actions (nodes *User Action Outcome* and *Agent Action Outcome* in Fig. 1). Agent actions are represented as decision variables in the framework, indicating points where the agent decides how to intervene. The fit of events with user's goals is modeled by the nodes class *Goals Satisfied*, which in turn influences the user's *Emotional States* (we call this part of the model the *appraisal-subnetwork*). Assessing user goals is not trivial, especially if asking the user about them during interaction is too intrusive, as is the case during game playing. Thus, our DDN also includes nodes (the *goal-assessment subnetwork*) to infer user goals from their interaction patterns and relevant traits (e.g., personality).

The subnetwork below the nodes *Emotional States* is the model's diagnostic part, representing the interaction between emotional states and their observable effects. *Emotional States* directly influence user *Bodily Expressions*, which in turn affect the output of *Sensors* that can detect them. Our framework is designed to modularly combine data from any available sensor, and gracefully degrade in the presence of partial or noisy information. It should be noted that the only temporal dependencies explicitly represented in Fig. 1 are between emotion variables, to account for the impact of the emotional state at time t on the emotional state at time $t+1$ (representing, for instance, the fact that the negative impact of a mismatched goal on one's emotion also depends on the preexisting emotional state). Other temporal dependencies may be relevant (e.g., between goals, as we discuss in Conati & Maclaren, 2009a), but require extra complexity to be captured reliably. Their absence in Fig. 1 should be seen as simplifying assumptions to be revised if empirical evaluations show a need for it.

Going from the high-level framework described here to concrete user models obviously requires filling in a large amount of often nontrivial details. In the rest of the chapter, we illustrate the process in the context of building an affective user model for an edu-game on number factorization, described next.

The Prime Climb Educational Game

In Prime Climb, students in sixth and seventh grade practice number factorization by pairing up to climb a series of mountains. Each mountain is divided into numbered sectors (see Fig. 2), and players must try to move to numbers that do not

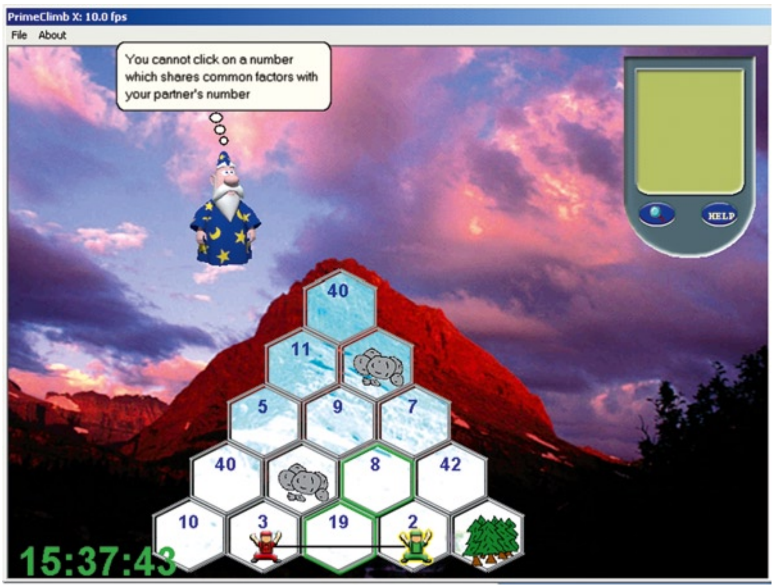

Fig. 2 The Prime Climb interface

share common factors with their partner's number, otherwise they fall. To help students, Prime Climb includes the Magnifying Glass, a tool that allows players to view the factorization for any number on a mountain in the PDA device displayed at the top-right corner on the game interface (see Fig. 2). Each student also has a pedagogical agent (Fig. 2) that provides individualized support, both on demand and unsolicited, when the student does not seem to be learning from the game.

When providing unsolicited hints, the agent currently decides when and how to intervene based solely on a probabilistic model that assesses how the player's factorization knowledge evolves during game playing (*learning model* from now on, described in (Manske & Conati, 2005)). The agent's interventions are structured as hints given at incremental levels, with the goal of triggering student reasoning about number factorization as they play (Conati & Manske, 2009).

- The first (*focus*) level aims to channel the student's attention on the skill that requires help. For instance, the agent says "Think about how to factorize the number you clicked on" if the student model predicts that the student does not know how to factorize that number.
- The second (*tool*) level is a hint that encourages the student to use the magnifying glass to see relevant factorizations.
- The third (*bottom-out*) level gives either the factorization of a number or which factors are in common between two numbers.

Students can choose to progress through the various levels by asking for further help. Otherwise, the agent goes through the progression when it needs to intervene on the same skill more than once. The above hints are provided regardless of the correctness of the student's move, if the learning model assesses that the student needs help with the relevant number factorization skills.

The affective user model described in the next section is designed to capture the affective reactions elicited in the student by his or her interaction with the game and the agent. It will eventually be integrated with the learning model to make agent interventions and game dynamics sensitive to both cognitive and affective states of the user.

Building the Affective Model

In this section, we illustrate the general steps needed to apply the framework described in the previous section to a specific learning environment (LE). For each step, we also discuss how the framework was applied to build the affective model for the Prime Climb game. We divide the description in two subsections, one for the causal and one for the diagnostic part of the model, since these two components are conceptually separate and could be adopted in isolation if desired, as it was done, for instance, in Conati and Maclaren (2009a).

Defining the Causal Component of the Affective Model

Define which emotions should be modeled. The OCC theory defines 22 emotions starting from the appraisal mechanism described in the previous section. These emotions include:

- Reactions related to how an occurring event impacts one's goals (*joy/distress* toward the event, *admiration/reproach* if the event was generated by a third party, *pride/shame* if was generated by oneself)
- Reactions about how an event impacts others that one may *like/dislike* (*happy-for/resentment* if the impact is positive, *pity/gloating* if the impact is negative)
- Emotions related to the prospective effects of an event (e.g., *hope/fear, relief/disappointment*)

Clearly, not all 22 emotions are always relevant for specific LEs. For instance, a one-user LE cannot elicit emotions related to other users. Even when a specific emotion is potentially relevant, inclusion in the model is a tradeoff between its impact on the interaction and the cost of modeling the dynamics that bring that emotion to bear. For instance, in Prime Climb we currently model 6 of the 22 OCC emotions: emotions toward game states (*joy/regret*), and related emotions toward the agent (*admiration/reproach* in the OCC theory) or toward oneself (*pride/shame*). The first four emotions were often informally observed during interaction with Prime Climb, and have been consistently self-reported by students during a variety of studies (Conati & Maclaren, 2009a). The *pride/shame* pair is clearly relevant to any kind of reward-based interaction, however, we have no formal evidence on the extent of its occurrence because of difficulties in obtaining reliable self-reports 'more on this in a later section'. Still, once the model is set up to capture *admiration/reproach*, adding *pride/shame* has little overhead because the only additional factor that needs to be tracked to distinguish between these two emotion pairs is whether the Prime Climb state currently appraised has been generated by the Prime Climb agent or by the student. In contrast, while we have substantial evidence that emotions toward the climbing partner (i.e., another student) arise frequently during game play, they are currently not included in the model because of the added complexity involved in modeling a two-player interaction. Because of this complexity, we decided to first evaluate the feasibility of the approach with the simpler model described here.

Define student goals. This step requires us to define the set of goals that students may have when using the target LE. These goals can either be well-defined objectives set by the game itself (*fixed goals* from now on) or more *subjective goals* still influenced by the type of interaction that the LE supports but not as obviously related to it as fixed goals are. While fixed goals can be easily defined from an analysis of the LE, the relevant set of subjective goals must be derived empirically by observing actual student interactions. For instance, observations and interviews of students playing Prime Climb uncovered six high-level non-mutually exclusive goals (*Have Fun, Avoid Falling, Beat Partner, Learn Math, Succeed By Myself and Wanting Help*). While some of these goals naturally derive from to the structure of

Prime Climb (*Have Fun, Avoid Falling, Learn Math*) others are more arbitrary. For instance, the goal *Beat Partner* is actually in contrast to the nature of Prime Climb, since the two players are supposed to collaborate, when climbing. The goals *Succeed By Myself* and *Wanting Help* intuitively seem mutually exclusive; however, we observed that they can in fact co-exist for students who express a general preference to succeed by themselves but end up wanting help during especially challenging episodes. We have collected the data to instantiate goal-related variables and their prior probabilities in the Prime Climb model via user studies in which students were given a post-game questionnaire to assess which of the above goals they experienced during game playing. The post-questionnaire includes goal-related statements to be ranked on a Likert scale (1–5), and there are multiple statements per goal, to increase the reliability of the students' answers (e.g., "I wanted to learn math by playing the game," "I didn't want to think about math when I was playing the game"). The questionnaire also includes an open-ended question gauging the presence of any additional goal, but none were found.

Define means for goal assessment. The goal set defined in the previous step specifies the range of goals each student *may* have while interacting with the target LE, not which goals the student actually pursues at any given point in time. So, unless goals are specifically set by the LE during interaction, they need to be inferred. One option is to endow the interface with an unobtrusive way for students to specify their goals while playing. Alternatively, the system needs to perform *goal recognition*, i.e., infer the goals dynamically as the student interacts with the system. In Prime Climb, we adopted the second approach, as eliciting student goals explicitly during game playing was deemed too intrusive. In particular, we leverage the fact that user goals are influenced by *user personality* (Costa & McCrae, 1992) and affect user *interaction patterns*, which in turn can be inferred by observing the outcomes of individual user actions. Thus, observations of both the relevant user traits and action outcomes can provide the DDN with indirect evidence for assessing user goals.

We derived the data to build the portion of the DDN that exploits this evidence for goal assessment via a series of Wizard of Oz studies where pairs of students interacted with the game while an experimenter controlled the pedagogical agent (Zhou & Conati, 2003). Students reported their goals via the questionnaire described above. Information on student personality is included in the model based on the Five-Factor Model (Costa & McCrae, 1992), which represents personality as five domains – *neuroticism*, *extraversion*, *openness*, *agreeableness* and *conscientiousness*. Data to instantiate the prior and conditional probabilities for the variables that represent these domains in the model was collected through a standard personality test (Graziano, Jensen-Campbell, & Finch, 1997). Interaction logs were mined to define the relationships between student goals (assessed via the goal post-questionnaire) and interaction behaviors. This process resulted in the goal assessment subnetwork shown in Fig. 3, where all nodes are binary variables. More details on the construction of this part of the model can be found in Zhou and Conati (2003).

Define appraisal relationships. Following the OCC appraisal model, a student's emotional state depends on whether his or her goals are satisfied or not during the interaction with a LE. Modeling this process in a DDN requires identifying how

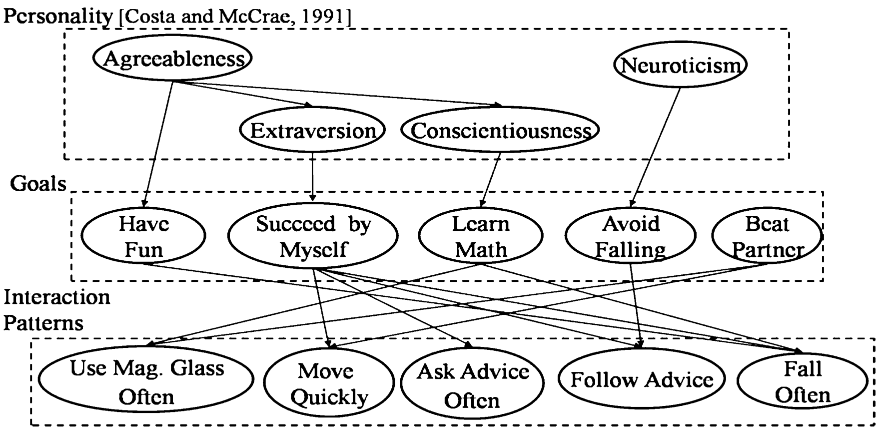

Fig. 3 Goal assessment portion of the model

each relevant game state relates to the set of possible student goals in terms of appraisal. Some of these relations can be defined intuitively. In Prime Climb, for instance, if the student has the goal *Avoid Falling*, a successful move likely satisfies it, while a fall likely does not. If the student has the goal *Beat Partner*, only a move that brings the player ahead of the partner on the mountain is likely to satisfy this goal. Other appraisal relationships must be derived empirically. For instance, we could not define a priori which events satisfy student goals *Have Fun* or *Learn Math*. Similarly, although an unsolicited hint from the Prime Climb agent intuitively violates the goal *Succeed-by-myself* and satisfies *Want-Help*, it is unclear how the various types of hints are appraised with respect to these goals given that they vary substantially in the amount of help that they provide. We defined the appraisal relationships in the Prime Climb affective model through a user study in which students, after game playing, were asked to rate propositions of the type "I <*goal-related action*> when <*game event*>." In each proposition, <*goal-related action*> is a statement related to one of target appraisal goals (e.g., "learnt math," "had fun") and <*game event*> is a relevant event in the Prime Climb interaction (e.g., "I fell," "the agent suggested to use the magnifying glass"). Figure 4 shows the appraisal relations derived from this process with respect to the outcome of student actions. More details about the process and the resulting model can be found in (Conati & Maclaren, 2009a).

Defining the Diagnostic Component of the Affective Model

Define sensors for the diagnostic part of the model. The choice of sensors to be included in the model largely depends upon the type of emotional states that the model must capture. There has been considerable success in linking individual

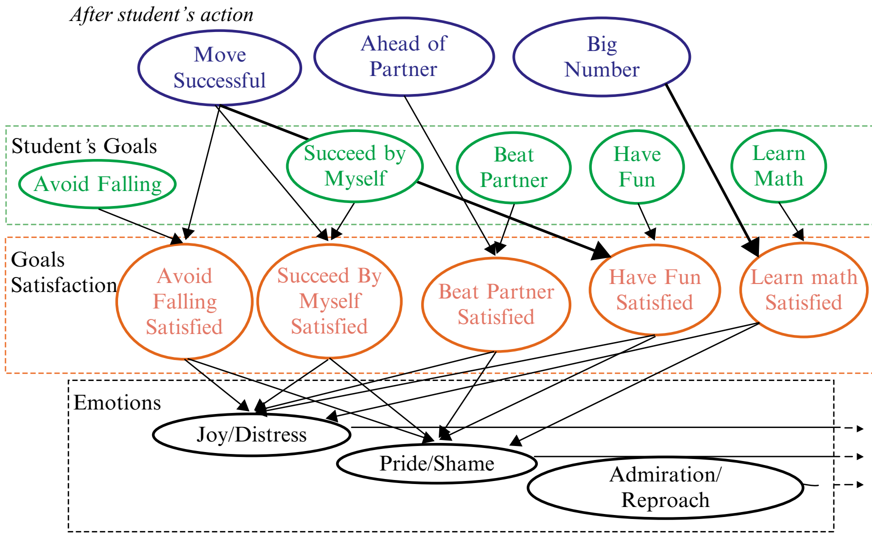

Fig. 4 Sample excerpt from the Prime Climb appraisal network

bodily/physiological expressions to the affective dimensions of valence and arousal, such as heart beat and measures of skin conductance (e.g., Prendinger et al., 2005), various facial expressions (e.g., Lang, Greenwald, Bradley, & Hamm, 1993), acoustic–prosodic and lexical speech features (Litman & Forbes-Riley, 2004). There have also been results on combinations of sensors as detectors of specific emotions. For instance, Healey and Picard (2005) report 89% accuracy in recognizing four levels of driver anxiety by integrating measurements from five physiological sensors, three video-cameras and a microphone. Cooper et al. (2010) linked measurements from a mouse that captures pressure placed on its various points, as well as camera-detected facial expressions with high student interest during interaction with an intelligent LE for math. They also linked facial expressions with high levels of student excitement. D'Mello and Graesser (2010) found that facial expressions as coded by external judges can discriminate among student states of confusion, boredom, frustration, and neutral. Although existing results can help guide the choice of sensors, the final selection should always be empirically validated, as sensors performance highly depends upon a variety of factors such as whether the emotions are spontaneous or artificially elicited, age of participants, and the interaction context.

In our research, empirical evaluations of the causal part of the Prime Climb model (Conati & Maclaren, 2009a) suggested that its performance could be improved by including information on the valence of the student affective states, leading us to experimenting with a sensor that would serve this purpose. More specifically, the causal model proved to be unable to reliably capture *regret* toward the agent because it could not capture the shifts that some students experience between the goals *Succeed-by-myself* and *Wanting Help* at critical times of game playing. This confusion causes the model to misjudge how students react to the agent's interventions (or lack thereof) at those times. Accurate goal recognition can be extremely challenging, but this particular problem could be alleviated if the Prime Climb model can detect when the student moves to a state of negative valence after an agent action. A sensor that has shown to be a good detector of negative affect is the electromyography (EMG) placed on the corrugators muscle on the forehead (Lang, Greenwald, Bradley, & Hamm, 1993). EMG sensors measure muscle activity by detecting surface voltages that occur when a muscle is contracted. When placed on the corrugator muscle on the forehead, the signal gets excited by this muscle's movements, and previous studies linked greater EMG activity in this area with expressions of negative affect.

Adding sensors to the model. The bottom part of Fig. 1 shows the most complete incarnation of this step, where the connection between affective states and sensors predictions is defined through the bodily expression that each sensor captures. Having the connection between affective states and sensors go through bodily expressions is advisable when using multiple sensors to detect a specific bodily expression (e.g., a videocamera and an EMG to detect eyebrow movements). This configuration requires the specification of both the conditional probabilities that express each sensor's reliability in detecting the target bodily expression, as well the conditional probabilities that encode the reliability of that bodily expression as an indicator of the target emotional state. Alternatively, sensor measurements can be directly linked to the target emotional state, as we did in our first exploration of the EMG sensor for the Prime Climb model.

Since we wanted the EMG sensor to provide information on affective valence, two new nodes were added to each time slice of the affective model: *Valence* and *Signal Prediction* (see Fig. 5, left), both binary. The *Valence* node represents the model's overall prediction for the student's affective valence the *Signal Prediction* node encodes whether the EMG signal predicts positive/negative valence at a time of interest. The conditional probability table (CPT) for *Valence* given *Emotional States* is defined so that the probability that valence is positive/negative is proportional to the number of positive/negative emotion nodes. The CPT for *Signal Prediction* given *Valence* represents the probability of observing an EMG prediction of positive or negative valence, given the student's actual affective valence. To instantiate this CPT, we ran a user study to collect both EMG evidence and accompanying affective labels.

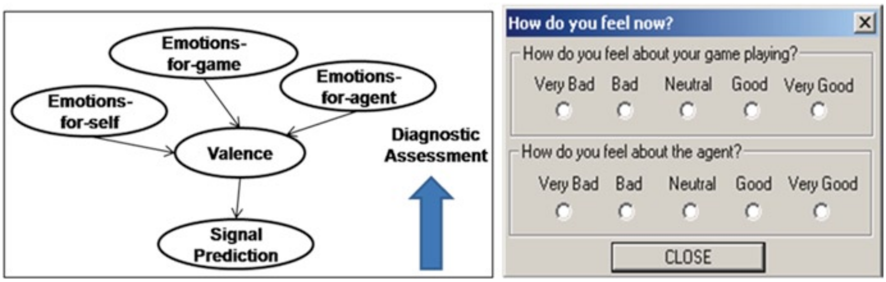

Fig. 5 Adding electromyography (EMG) data to the model (*left*); emotion self-report box (*right*)

The study involved 41 students (sixth and seventh grade) and its design was similar to the studies we used to instantiate other parts of the model. Here, however, each student had an EMG sensor placed on the forehead. During game play, students periodically self-reported their emotions via the dialog box shown in Fig. 5, a self-reporting mechanism that we have extensively validated and used throughout this research (Conati, 2004; Conati & Maclaren, 2009a).[1]

The log files from the study include all relevant game events (e.g., a student's successful climbs and falls, agent interventions), the student's reported emotions and the EMG signals sampled at 32 Hz. These log files were analyzed to generate a set of datapoints of the form <*affective valence, signal prediction*>, where a datapoint is created for each logged event that can be associated with an emotion self-report. The value for *affective valence* (positive or negative) is derived from that self-report; the value for *signal prediction* (also positive or negative) is computed by analyzing the EMG signal in the 4 seconds following the event. The analysis yielded 196 datapoints, which were used to instantiate the CPT for the *Signal Prediction* node in Fig. 5 by calculating the frequencies of the various combinations of *signal prediction/affective valence* value pairs in the data set. More details on this process can be found in (Conati & Maclaren, 2009b).

Model Evaluation

The data from the study described in the previous section was used to evaluate the resulting Prime Climb affective model with respect to two main questions

1. Is the goal assessment mechanism sufficiently accurate to support our appraisal-based modeling approach?
2. How does the model using only causal information compare to the model that includes diagnostic information from the EMG sensor?

The general evaluation methodology is to compare various versions of the model by using a Prime Climb simulator. The simulator is used to feed log files from the study to each model that is to be evaluated. Model predictions of affect are collected at points in which students generated their emotion self-reports, and compared with the reported emotions (in this study, 170 reports of Joy, 14 reports of Distress, 127 reports of Admiration, and 28 reports of Distress).

To answer question 1, we compared the performance of the causal model with the goals assessed via the mechanism described earlier against a model where goals were set based on student post-questionnaire responses from the study. The model's performance when using goal assessment increased significantly for *Distress* and

[1] Currently, the dialog box only elicits information on emotions towards the game and the agent because dealing with three pairs of emotions turned out to be too overwhelming for students.

for *Reproach*, mostly because dynamic goal assessment can capture to some extent the changes in student goals during interaction, which the alternative model cannot do since goal values as set up-front. The performance of the model with goal assessment for *Joy* and *Admiration* decreased slightly (from 69.6 to 68.7% for *Joy*, from 67.2 to 66% for *Admiration*) but the decrease is not statistically significant. Thus, from a practical standpoint, the model using the probabilistic goal assessment performs better than the model using explicit evidence on student's goals. The only way to further improve goal information in the model would be to obtain self-reports on students' goals periodically during interaction.

To answer question 2, we compared the performance of the causal model against the performance of the model that includes EMG. We found that, for datapoints corresponding to strong, consistent emotional states (e.g., when students reported emotions with the same valence toward the game and the agent) the complete model performs significantly better than the predictive model. Accuracy for reproach went from 39 to 63%, bringing overall accuracy on emotions toward the agent to 73% (against 61% for the causal mode). Accuracy for emotions toward the game went from 72.6 to 76.9%, also a statistically significant increase. In contrast, the addition of the EMG sensor made no difference for data points corresponding to weaker or conflicting emotional states (e.g., states in which students reported at least one emotion to be neutral, or emotions with opposite valence toward game and agent). In the presence of weak emotions, likely the affective reaction is not strong enough to generate movements of the corrugators muscle that are detectable by the EMG sensor. Thus, adding to the model more sensitive sensors for valence detection may alleviate this problem. In the case of conflicting emotions, the addition of the EMG brings no value because it captures overall valence but does not help discriminate valence at the level of the individual emotion pairs. This problem calls for refinement in the goal assessment process, to better capture shifts between goals or goal priority during interaction.

Conclusions

We have presented an approach to modeling user affect that combines explicit information on both causes and effects of emotional reaction. One advantage of this approach is that using both 'of these' sources increases model accuracy. A second advantage is that, by assessing not only which emotions the student is feeling but also *why* they arise, this model enhances a LE's ability to adequately respond to these emotions. For instance, if the LE can recognize that the user feels a negative emotion because of something wrong the user has done (*shame* in our models) it may provide hints aimed at making the user feel better toward herself. If the LE recognizes that the user is upset because of its own behavior (*reproach* in our models), it may take actions to make amends. These specific interventions are more difficult to identify with approaches that do not have such an explicit representation of the reasons underlying user emotions (e.g., Cooper et al., 2010; D'Mello &

Graesser, 2010). This added value, however, comes with increased model complexity. Implementing the appraisal mechanism that enables causal assessment requires defining relationships between student traits, goals, and events. This process often involves laborious data collection, as we illustrated in this chapter, with our experience in building the affective model for the Prime Climb edu-game. It is our long-term objective to compare the approach presented here with lighter-weight models, to better understand if and when the added cost is worth the effort.

A more immediate goal is to integrate the predictions of the Prime Climb affective model with the existing model of student learning, so that game dynamics and agent interventions can be tailored to both. Toward this end, we are conducting user studies to understand specific limitations of agent hints based solely on the learning model, and how affect-sensitive responses may overcome these limitations. We are also exploring ways to elicit explicit information on student goals at selected times during interaction in order to better cope with situations in which the model cannot reliably assess these goals. The objective here is to maximize the value of this information for the model, without excessive disruption to game play. Similarly, we want to investigate if and when it is appropriate to explicitly ask students about their emotions, to cope with situations in which the model does not have sufficient information to generate a confident assessment.

References

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence, 16*(7–8), 555–575.

Conati, C. (2004). How to evaluate models of user affect? In In Elisabeth André, Laila Dybkjær, Wolfgang Minker, & Paul Heisterkamp (Eds.),*Proceedings of ADS'04, Tutorial and Research Workshop on Affective Dialog Systems*. (pp. 288–300), Kloster Irsee, Germany.

Conati, C., & Maclaren, H. (2009a). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267–303.

Conati, C., & Maclaren, H. (2009b). Modeling user affect from causes and effects. In *Proceedings of UMAP 2009, First and Seventeenth International Conference on User Modeling, Adaptation and Personalization.* pp 4–15. Springer, LNCS 5535.

Conati, C., & Manske, M. (2009). Evaluating adaptive feedback in an educational computer game. In Hannes Högni Viljálmsson (Eds.), *Proceedings of Intelligent Virtual Agents (IVA) 2009* (pp. 146–158). Amsterdam, The Netherlands: Springer, LNCS 5773.

Cooper, D., Muldner, K., Arroyo, I., Park Woolf, B., & Burleson, W. (2010). Ranking feature sets for emotion models used in classroom based intelligent tutoring systems. In P. De Bra, A. Kobsa, & D. Chin (Eds.), *Proceedings of UMAP 2010* (pp. 135–146). Big Island, HI, USA: Springer, LNCS 6075.

Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*, 653–665.

D'Mello, S., & Graesser, A. C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*(2), 147–187.

de Castell, S., & Jenson, J. (2007). Digital games for education: When meanings play. *Intermedialities, 9*, 45–54.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Graziano, W. G., Jensen-Campbell, L. A., & Finch, J. F. (1997). The self as a mediator between personality and adjustment. *Journal of Personality and Social Psychology, 73*, 392–404.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156–166.

Lang, P., Greenwald, M., Bradley, M., & Hamm, A. (1993). Look at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261–273.

Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer-human tutoring dialogues. In Donia Scott (Ed.), *42nd Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 352–359). Barcelona, Spain.

Manske, M., & Conati, C. (2005). Modelling learning in educational games. *Proceedings of the 12th International Conference on AI in Education*. Amsterdam, The Netherlands.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotion*. New York: Cambridge University Press.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Recognizing, modeling, and responding to users' affective states. In Liliana Ardissono, Paul Brna & Antonija Mitrovic (Eds.), *UM'05 10th International Conference on User Modeling*. (pp. 60–69). Edinburgh, UK. Springer, LNCS 3538.

Zhou, X., & Conati, C. (2003). Inferring user goals from personality and behavior in a causal model of user affect. In Johnson, Lewis & Andre, Elisabeth (Eds.), *IUI'03, International Conference on Intelligent User Interfaces* (pp. 211–218). Miami, FL.