# Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions

Christos N. Moridis, *Member*, *IEEE*, and Anastasios A. Economides, *Senior Member*, *IEEE*

**Abstract**—Empathetic behavior has been suggested to be one effective way for Embodied Conversational Agents (ECAs) to provide feedback to learners' emotions. An issue that has been raised is the effective integration of parallel and reactive empathy. The aim of this study is to examine the impact of ECAs' emotional facial and tone of voice expressions combined with empathetic verbal behavior when displayed as feedback to students' fear, sad, and happy emotions in the context of a self-assessment test. Three identical female agents were used for this experiment: 1) an ECA performing parallel empathy combined with neutral emotional expressions, 2) an ECA performing parallel empathy displaying emotional expressions that were relevant to the emotional state of the student, and 3) an ECA performing parallel empathy by displaying relevant emotional expressions followed by emotional expressions of reactive empathy with the goal of altering the student's emotional state. Results indicate that an agent performing parallel empathy displaying emotional expressions relevant to the emotional state of the student may cause this emotion to persist. Moreover, the agent performing parallel and then reactive empathy appeared to be effective in altering an emotional state of fear to a neutral one.

**Index Terms**—Computers and education, intelligent agents, empathy, user interfaces

## 1 Introduction

A main focus of research relating to any kind of interactive computerized environment, ranging from video games to tutoring systems, has to do with Embodied Conversational Agents (ECAs) and Avatars. ECAs are digital models determined by computer algorithms, whereas avatars are digital models guided by humans in real time [1]. Thus, avatars' interaction is human controlled, whereas ECAs have an automated, predefined behavior. While this paper focuses on ECAs empathetic behavior in the context of a self-assessment test, some of the conclusions of this work could be interesting for applications using avatars as well.

Regarding ECAs, Cassell and Miller [2] have suggested that an ECA: "... must be capable of simulating many of the same responses that humans give, such as happiness and sadness, attentiveness and boredom, desire to take the floor, and acknowledgment that the other's words have been understood." Moreover, an ECA developed for a tutoring system should know when and how to intervene in order to influence the student's emotional state, based on an educational pedagogy integrating emotional models in learning [3].

Empathetic behavior has been suggested to be one effective way for ECAs to provide feedback to learners' emotions. Carl Rogers [4] defined empathy as the ability to perceive another person's inner psychological frame of report with precision, but without ever losing consciousness of the fact that it is a hypothetical situation. Therefore, empathy is to feel, for example, someone else's pain or pleasure and to perceive the cause of these feelings as perceived by the other person, without setting aside self-awareness. Importantly, there is research evidence indicating that humans interact with computers in a way similar to the social behavior exhibited in human-human interactions [5], [6]. Furthermore, a number of authors have argued that the presence of empathic emotion in a computer agent has significant positive effects on a user's impression of that agent and, as a result, will advance human-computer interaction [7], [8], [3].

Moreover, an issue that has been raised is the effective integration of parallel and reactive empathy into ECAs involved in tutoring systems. Parallel empathy describes a person displaying an emotional state alike to that of another individual. This is usually based on a considerate attitude toward another's individual emotional state and expresses a person's ability to identify with the emotions of that individual [9]. Reactive empathy aims at another's individual emotional state, trying to provide insight for recovering from that state. Thus, reactive empathy may involve a person displaying emotions that are different from those of his/her interlocutor so as to change the other individual's affective state [9].

In this study, three identical female ECAs with three different types of empathetic behavior were displayed as feedback to students' Happy, Sad, and Fear emotional states in the context of a self-assessment test. Then their affective transitions from these three emotions to Happy, Angry, Sad, Surprised, Scared, Disgusted, plus Neutral emotional states were examined.

• *The authors are with the Department of Information Systems, University of Macedonia, 156 Egnatia Avenue, Thessaloniki, Hellas 54006, Greece. E-mail: {chmoris, economid}@uom.gr.*

*Manuscript received 8 Nov. 2010; revised 15 Dec. 2011; accepted 2 Mar. 2012; published online 20 Mar. 2012.*
*Recommended for acceptance by C. Conati.*
*For information on obtaining reprints of this article, please send e-mail to: taffc@computer.org, and reference IEEECS Log Number TAFFC-2010-11-0107.*
*Digital Object Identifier no. 10.1109/T-AFFC.2012.6.*

A lot of research still needs to be conducted to arrive at conclusive evidence concerning different forms of ECAs' empathetic behaviors displayed under different learning contexts. Accordingly, few studies have been done with the purpose of finding appropriate expressions for empathetic ECAs displayed in tutoring systems, while none of them has specifically examined empathetic agents' verbal behavior combined with nonverbal communicational channels. The present study is a first step toward this direction. It aims to examine the impact of ECAs' emotional facial and tone of voice expressions combined with empathetic verbal behavior, when displayed as feedback to student's fear, sad, and happy emotions in the context of a self-assessment test.

Students' emotional experiences while using a self-assessment test system may have some particularities which could also involve "basic" emotions. When the effect of negative emotions, such as Sad and Fear, is too intense, the student's performance can be seriously impaired. Frequent errors could create the expectation of more errors, thus increasing negative emotions and leading to even more wrong answers until the student's performance collapses [10]. Fear of failure has been stated to be an important factor during test taking. As shown in [11], some students' self-defeating beliefs and fear of failure had a strong association with eventual test failure, the very situation that they were trying to avoid. Positive emotions may also occasionally necessitate instruction. For instance, providing the correct answer to a hard question could induce positive emotions such as joy and enthusiasm, but also lead to loss of concentration if too much consideration is given to the elicited emotions. With no pedagogical feedback, positive emotions can lead students focus on excitement and undervalue the effort required to achieve a successful result [12], [13]. Moreover, affective feedback aimed at corresponding to the needs of personalized self-assessment should be adaptable even when learners experience uncommon emotional states.

In order to develop affective tutoring systems, it is essential that the ways in which emotions influence learning be known. Nevertheless, this knowledge would have no use in affective tutoring systems if these systems could not recognize a student's emotional state [3]. Preferably, data from many modes of interaction should be combined by a computer system so that it can make as valid estimations as possible about users' emotions [14], [15].

In this study, students' emotions during a computerized self-assessment test were observed by two independent researchers and the FaceReader. The FaceReader, developed by Vicar Vision and Noldus Information Technology, recognizes facial expressions by distinguishing six basic emotions (Happy, Angry, Sad, Surprised, Scared, Disgusted, plus Neutral) with an accuracy of 89 percent [16]. The system is based on Ekman and Friesen's theory of the Facial Action Coding System (FACS) that states that the basic emotions correspond with facial models [17].

Feedback was displayed only when both the FaceReader and the researchers agreed that a student was in a Fear, Sad, or Happy emotional state. All emotions during the test, including the affective transitions observed as a result of each feedback, were registered only when both the FaceReader and the researchers agreed that a student was in a Neutral, Sad, Happy, Surprise, Disgust, Fear, or Angry emotional state.

Students participating in the self-assessment test were randomly distributed into four groups:

1. a group where no feedback was displayed (N.F. group),
2. a feedback group where an ECA performing parallel empathetic behavior with a neutral facial and tone of voice expression was displayed (ECA 1 group),
3. a feedback group where an ECA was performing parallel empathetic behavior, displaying a facial and tone of voice expression that was relevant to the emotional state of the student (Fear, Sad, Happy) (ECA 2 group), and
4. a feedback group where an ECA was performing parallel and then reactive empathetic behavior, displaying a facial and tone of voice expression that was relevant to the emotional state of the student (Fear, Sad, Happy) for parallel empathy and then displaying an emotional facial and tone of voice expression different from the emotional state of the student for reactive empathy (ECA three group).

To summarize, this paper examines the inclusion of emotional facial and tone of voice expressions in the delivery of parallel and reactive empathy, and its impact on how students' emotions of Happiness, Fear, and Sadness were altered as a result of this kind of agents' empathetic behavior.

The paper continues as follows: Section 2 briefly describes some research paradigms of related work. Section 3 refers to the methodology, as related to issues concerning the design of the ECAs used in this study and issues concerning the design of this study's experimental environment. Section 4 is a description of the experimental process. Section 5 describes the methods used for analyzing the experimental data. All results are presented in Section 6. Section 7 is a conclusion section, discussing significant findings of this work and issues concerning future research.

## 2 Related Research

Early pioneers of affective tutoring systems have stressed the importance of developing mechanisms that will render these systems capable of recognizing and reacting to students' affective states [18], [19]. For instance, Conati's [18] work on probabilistic assessment of affect was based on the Ortony, Clore, and Colins (OCC) model [20], combining data from situational appraisals that generate emotions with bodily reactions associated with emotion appearance. Thus, the predictions based on the OCC model are confirmed from a learner's bodily expressions to increase the recognition's accuracy. In that sense, an automatic affect recognizer would enable the tutoring system to intervene appropriately to a learner's affective states (for a comprehensive review of other affective models see the work of Calvo and D'Mello [21]).

Although the OCC theory defines 22 different emotions, affective learning researchers have considered other emotions to be relevant to learning as well. Learning-centered emotions such as confusion, boredom, flow, and frustration, have been shown to be particularly related to users' affective experiences taking place during learning activities [22]. However, this paper is committed to addressing Happy, Fear, and Sad basic emotions in the context of a self-assessment test.

As already discussed (see Section 1), students' emotional experiences while using a self-assessment test system may have some particularities which could also involve basic emotions. Interestingly, Calvo and D'Mello [21] in their review about affect detection, methods, models, and relevant applications have made the point that "basic emotions have minimal relevance to learning sessions that span 30 minutes to 2 hours." On the other hand, a self-assessment test usually lasts a lot less than 30 minutes, though the one presented in this study lasted 45 minutes, increasing the possibility for users to experience more learning-centered emotions (e.g., boredom). Nevertheless, the focus of this study is still on basic emotions that occur during a self-assessment test and the test was chosen to last 45 minutes so as to have the chance to collect more data from each user.

However, research identifying learning-centered emotions, such as boredom and frustration, is not on the whole different from researching basic emotions during learning sessions. Woolf et al. [23] proposed that emotions in a learning context tend to differentiate regarding the kind of emotional incident experienced by students. That is to say, for example, that in an educational context, anger may include a cognitive component that may lead to frustration. To treat this issue, these researchers addressed emotions occurring in learning as "cognitive-affective" terms. Thus, according to this approach, each basic emotion can be seen on a scale. For instance, the proposed scale for fear is "I feel anxious ... I feel very confident." Experiments with as many emotional states as possible will provide valuable empirical knowledge.

It has been suggested that learners who enter an affective state detrimental to learning are likely to stay in that state, entering a "vicious cycle" that prevents them from actively reengaging in the constructivist learning process [24], [25]. Baker et al. [24] have suggested that considerable effort should be put into recognizing and giving adequate feedback to boredom and confusion, with an emphasis on developing pedagogical interventions to interrupt the "vicious cycles" occurring when a student becomes bored and remains bored.

Accordingly, it has been shown that empathetic feedback expressed through an agent can change the affective state of the learner [26], [27]. Moreover, there is research evidence supporting that students interacting with an empathetic agent would show higher self-efficacy and be more interested in learning tasks [28].

Although ECAs expressing empathetic behavior are increasingly gaining attention, there are still a few research paradigms attempting to analyze and examine the role of ECAs' empathetic behavior in a learning context. Some of these are presented in the following paragraphs.

An early approach was that of Lester et al. [29], implementing a life-like pedagogical agent named COSMO, aiming to provide tutoring help in a learning environment for the domain of Internet packet routing. COSMO does not resemble a virtual human, but rather an "impish, antenna-bearing creature." The agent is capable of performing full-body emotive behaviors in reply to learners' problem-solving actions. Moreover, the COSMO agent utilizes empathetic behavior as feedback to learners' experiences of disappointment and sadness.

Hone et al. [30] programmed an embodied character using Microsoft Agent and Visual Basic. Microsoft Agent is a collection of programs for Microsoft Windows featuring animated characters that are capable of talking, performing facial expressions, and bodily displays of emotion. A virtual human character, James the butler, was chosen from the Microsoft Agent character collection. Hone et al. integrated various strategies into the agent to reduce negative emotions. Moreover, some of the agent behaviors were derived from human displays of empathy. The agent was displayed in a biology learning environment for university undergraduates (18-25 years old). Results confirmed the agent's ability to reduce negative emotions (frustration, boredom, and depression).

Burleson and Picard [31] developed an Affective Agent Research Platform consisting of a character agent able to perform a wide variety of expressive interactions. The agent resembles a "humanoid robot" and is capable of mirroring a number of nonverbal behaviors supposed to influence persuasion, liking, and social rapport. Furthermore, the agent responds to frustration with empathetic or task-support dialogue. Burleson and Picard examined the impact of this agent on 11-13-year-old children while helping them solve the Towers of Hanoi problem. Results from this study revealed gender-specific impacts of the agent's nonverbal behaviors and affective support strategies on children's frustration and perception of the agent.

Invaluable research has emanated from the Autotutor project. Autotutor is a more than 10-year-old project aiming at developing an intelligent tutoring system that enhances students' learning by maintaining a conversation in natural language [32], [33], [34], [35]. The agent appears as a virtual human and interacts with the learner through synthesized speech, facial expressions, and simple hand gestures. Empathetic behavior has also been integrated in the autotutor as feedback to learners' emotions of boredom and frustration [26]. Results from numerous experiments conducted with the AutoTutor "teaching" college students Newtonian physics, computer literacy, and scientific reasoning indicate that the system can greatly enhance learning gains.

Some recent research efforts toward empathetic ECAs have been conducted in the context of a narrative-centered inquiry-based learning environment, the CRYSTAL ISLAND [36]. There are six virtual human characters (Audrey, Elise, Jin, Quentin, Robert, and Teresa) in the CRYSTAL ISLAND environment, each one of them playing a distinct role. The environment was developed to provide tutoring for the fields of microbiology and genetics to middle school students. The empathetic characters respond to students' emotions (anger, anxiety, boredom, confusion, delight, excitement, fear, flow, frustration, and sadness), employing parallel and reactive empathy. Results of these experiments have been very promising, demonstrating the agents' capability to alter the students' emotional state [25], [27], [37].

TABLE 1
The ECAS' Synchronized Speech and Facial Expressions

| FEAR-ECA 1 | FEAR-ECA 2 | FEAR-ECA 3 |
|---|---|---|
| **Voice:** Somehow this test makes you feel afraid **Facial expression**: Neutral | **Voice:** Somehow this test makes you feel afraid. **Facial expression:** Fear | **Voice:** Somehow this test makes you feel afraid. Cheer up, continue trying and you will succeed. **Facial expression:** Fear and then happy |
| **SAD-ECA 1** | **SAD-ECA 2** | **SAD-ECA 3** |
| **Voice:** Somehow this test makes you feel sad **Facial expression**: Neutral | **Voice:** Somehow this test makes you feel sad **Facial expression:** Sad | **Voice:** Somehow this test makes you feel sad. Cheer up, continue trying and you will succeed. **Facial expression:** Sad and then happy |
| **HAPPY-ECA 1** | **HAPPY-ECA 2** | **HAPPY-ECA 3** |
| **Voice:** Somehow this test makes you feel happy. **Facial expression:** Neutral | **Voice:** Somehow this test makes you feel happy **Facial expression:** Happy | **Voice:** Somehow this test makes you feel happy. Continue the test with attention **Facial expression:** Happy and then Neutral |

In this paper, we followed the same approach concerning the parallel and reactive empathetic mechanisms performed through the ECAs as in [27], [36], and [37]. Nevertheless, none of the studies mentioned in this section specifically searched for the effects of empathetic behavior with nonverbal cues, such as facial expressions. The present study attempts to fill this void by examining the effects of ECAs' emotional facial and tone of voice expressions when combined with parallel and reactive empathy in the context of a self-assessment test.

## 3 Methodology

### 3.1 Design of the ECAs

Three identical female 3D ECAs with three different kinds of empathetic behavior were implemented for this experiment in order to be displayed as feedback to students' Happy, Sad, and Fear emotions during a self-assessment test. Each ECA displayed a different empathetic behavior for each of these three emotions. The ECAs' emotional facial expressions were accompanied by an equivalent emotional tone of voice during speech. The synchronized speech and facial expressions of the ECAs are shown in Table 1. The instances shown were given in another language and were translated into English for the purposes of this paper. Only one instance of the text (shown in Table 1) of feedback for each ECA/Emotion case was examined. We chose to do so in order to be as certain as possible about the effect of this particular combination.

The three ECAs behaved as follows:

- *ECA 1*: Displayed parallel empathetic behavior relevant to the student's emotion (Happy, Sad, or Fear), performing neutral facial and tone of voice expressions.
- *ECA 2*: Displayed parallel empathetic behavior relevant to the student's emotion (Happy, Sad, or Fear), displaying facial and tone of voice expressions that were relevant to the emotional state of the student.
- *ECA 3*: When the student was in a Sad or in a Fear emotion, ECA 3 displayed empathetically encouraging behavior, displaying facial and tone of voice expressions that were relevant to the emotional state of the student (Sad or Fear) synchronized with the parallel empathetic behavior, and then Happy facial and tone of voice expressions synchronized with the reactive empathetic behavior. When the student was in a Happy emotion, ECA 3 displayed empathetically encouraging behavior, displaying facial and tone of voice expressions that were relevant to the emotional state of the student (Happy) synchronized with the parallel empathetic behavior, and then Neutral facial and tone of voice expressions synchronized with the reactive empathetic behavior.

Based on Yee et al.'s [38] meta-analysis on embodied agents that stated that agents with higher realism are generally rated more positively than those with lower realism, we preferred to develop a virtual human rather than an animal character. We also preferred to develop a female agent, stimulated by Hone's [39] research evidence that female representations may be more effective at reducing frustration than male representations. Hone suggested that this may be attributable to the stereotype that females are usually more empathetic than males. Interestingly, gender stereotypes coming from the real world could apply to human-computer interaction [6], [40].

In their experiments, researchers have used several ways to express empathy through an embodied agent. In some cases they merely used specific sentences, such as "It seems you did not like this question so much" [39], [41]. In other cases, empathetic verbal behavior was synchronized with simple emotional expressions [42], such as joy or anger. In still other cases, researchers have even used complex emotional expressions, where different emotions can be expressed on different areas of the face [43]. However, to the best of our knowledge, there has been no previous effort to examine the synchronization of empathetic verbal behavior with emotional facial expressions in the context of a learning environment.

In this experiment, we used simple facial expressions (Happy, Fear, and Sad) to examine the influence of presence/absence of emotional facial expression (Fig. 1) of an empathetic agent on student's emotions. An identical set of behavior, displayed by identical virtual humans, with only one difference (in this case facial expression and tone of voice), could provide valuable information [2].

It was very crucial for this experiment to address whether each ECA's emotional facial expression can be assigned to the relevant emotion as well as if the participants were

Fig. 1. The ECA in Sad and Happy facial expressions.

capable of perceiving it as such. Since the participants of the self-assessment test did not come to the experiment all at once, we did not ask them to validate the ECAs' emotional expressions. We were afraid that if we did so, students who were about to participate in the experiment during the subsequent days would be biased by being informed by their colleagues that emotions did play a role in this experiment. Therefore, we chose to validate the ECAs emotional expressions only among 30 subjects (irrelevant to the self-assessment test experiment), recruited through advertisement, selected to be as close as possible to the demographic characteristics (race, income, age, educational attainment, location, and computer literacy) of the self-assessment test participants. Thus, the effect of each ECA had already been measured at a previous stage using the 30 (15 females and 15 males) subjects recruited through advertisement. Each one of the 30 subjects was asked to complete a questionnaire, composed of the images of all ECAs' facial expressions. More specifically, subjects were called to assign to each ECA's image an emotional state among Angry, Neutral, Sad, Happy, Disgust, Surprise, and Fear. Results indicated that Happy and Sad facial expressions were easily recognized by the participants with high percentages, 93 and 97 percent, respectively. Fear and Neutral facial expressions were recognized with lower percentages by the participants, 73 and 77 percent, respectively. Fear was mostly confused with Surprise and Neutral with Angry.

In our final experiment, however, facial expressions were accompanied by speech with a related tone of voice because we believed that this combination could improve recognition of the target emotions. The relevant sentences were uttered by a trained actress to convey the desired emotion. Although we didn't evaluate the effectiveness of the actress's speech, we trust that it was adequate given that the actress we hired was a professional.

### 3.2 Design of the Experimental Environment

An online multiple choice question test system, constructed for a previous experiment [44], [45], was adjusted to serve the needs of the current study. The system was developed within a Windows XP machine using JavaScript with Perl CGI on Apache web server with MySQL.

Throughout the test, a student selected his/her answer among four possible answers and confirmed his/her choice by clicking the "submit" button. After each question the system informed the student whether his/her answer was right or wrong and presented his/her score. The student could proceed to the next question by clicking the "next" button.

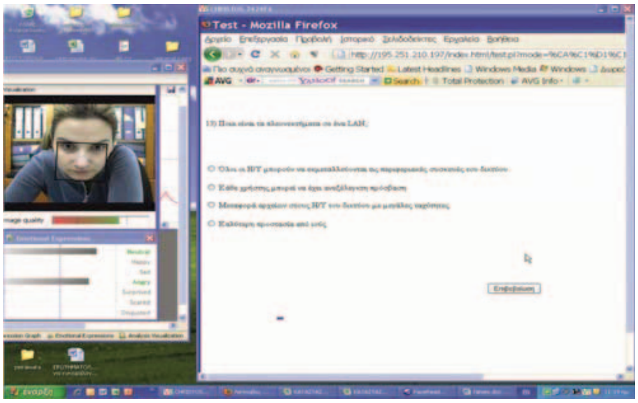

Fig. 2. Researchers' screen: FaceReader and VNC viewer (student's screen).

Each student took the test alone in an appropriately designed room. The room had two spaces, divided by a curtain. In the first space, there was the PC on which the self-assessment was administered. Moreover, the camera of the FaceReader was hidden in a bookcase. It is well known that people express themselves more freely when they feel that they are not being observed. The two researchers were in the second space. FaceReader was connected with another PC in that space, so the researchers were able to watch the facial expressions and the emotions of the participants in real time. The two researchers were also able to observe the student's actions during the test through VNC viewer software, which was presenting the student's screen in a separate window on the researchers' screen (Fig. 2). Each researcher recorded the student's emotions measured by the FaceReader and his/her estimation regarding the student's emotions at the same time, based on the student's facial expressions and actions.

This setting was chosen because emotion recognition through facial expressions during a computerized test is a very challenging task since students' facial expressions during a test have particularities that can mislead emotion recognition. For instance, many times FaceReader measured an angry emotion simultaneously with a neutral one, but neutral was the only emotion experienced by the students. This particular disagreement was expected. When participants read the questions, many of them had clouded eyebrows. People display this facial expression when reading something with great concentration [46].

Thus, the two researchers made their judgments based on the student's facial and bodily expressions (captured on camera), FaceReader's emotional recordings, and the student's interaction with the system (observed through VNC viewer). Those three sources of information enabled the researchers to independently assess the student's emotions. For instance, in a case where the researchers saw from the VNC viewer that the system had just informed the student that he/she had provided a correct answer to a question while at the same time FaceReader was recording a sad emotion, probably the researchers would not agree with FaceReader's "opinion" and thus feedback would not be triggered.

On the other hand, in case both FaceReader and the researchers agreed that a student was experiencing a sad emotion, but at the time the researchers were ready to trigger feedback FaceReader had started recording a

TABLE 2
Participants' Distribution at the Four Groups

| GENDER | GROUP | | | | TOTAL |
|---|---|---|---|---|---|
| | N.F. | ECA 1 | ECA 2 | ECA 3 | |
| MALE | 16 | 14 | 16 | 14 | 60 |
| FEMALE | 26 | 28 | 29 | 29 | 112 |
| TOTAL | 42 | 42 | 45 | 43 | 172 |

different emotion (e.g., happy), feedback would again not be displayed to the student. In cases like that, the researchers did not have the time to assess FaceReader's "opinion" in comparison to the other two sources of information; however FaceReader's "disagreement" was considered to be an adequate reason for not taking the risk of displaying an inappropriate feedback.

All emotions during the test, including the affective transitions observed as a result of each feedback, were registered only when both the FaceReader and the researchers agreed that a student was in a Neutral, Sad, Happy, Surprise, Disgust, Fear, or Angry emotional state.

The experiment was a Wizard of Oz [2] in which feedback was triggered manually only when both the FaceReader and the researchers recorded a Happy, Sad, or Fear emotion.

## 4 Experimental Process

### 4.1 Participants

Participants were first-year undergraduate students. The course was a basic IT (Information Technology) skills course and the syllabus included IT knowledge and techniques. Students were told that they could optionally participate in a self-assessment multiple choice questions test to help them assess their knowledge prior to exams. Those who wished to participate completed an application form in order to arrange an appointment. Two hundred eight applications were collected. The next step was the arrangement of the appointments. Since the purpose of the self-assessment test was to help the students assess their knowledge prior to exams, it was left to students to decide when they would feel that such a test would be helpful to them. Eventually, 172 applicants out of the 208 appeared at their appointments. The average age of students was 18.4 ($SD = 1.01$). From the 172 students, 60 were male (35 percent) and 112 were female (65 percent).

### 4.2 Material

The multiple choice questions were focused on course material taught in lectures. The content of the questions was prespecified by the course instructors prior to the study. The test consisted of 45 questions. The order of questions presented was randomly altered among students.

### 4.3 Procedure and Data Collection Methodology

The duration of the test was approximately 45 minutes. Each participant was randomly assigned to one of the four groups (Table 2):

1. Non-Feedback (N.F.) group,
2. ECA 1 group,
3. ECA 2 group, and
4. ECA 3 group.

Thus, 42 students were assigned to the N.F. group (16 males and 26 females), 42 students were assigned to the ECA 1 group (14 males and 28 females), 45 students were assigned to the ECA 2 group (16 males and 29 females), and 43 students were assigned to the ECA 3 group (14 males and 29 females).

During the test, one of the two independent researchers and the FaceReader were recording the participant's emotions. Because of the large number of participants, the two researchers were alternately observing the participants' emotions, so for each participant one of the two researchers was behind the curtain.

Facial expressions are often a mixture of emotions, so two (or even more) emotions may occur simultaneously. FaceReader's facial configurations correspond to multiple prototypical expressions. FaceReader first builds a model of the face and then classifies this model using classifiers trained on large sets of prototypical expressions of basic emotions. Facial Action Coding System (FACS) analysis is time consuming and it requires very intensive training. It is, consequently, not appropriate for large data sets. FaceReader is capable of analyzing facial expressions real time and is thus an interesting choice. The actual classification of the facial expressions by FaceReader is achieved by training an artificial neural network. Almost 2,000 manually annotated images were used as training data. Further details of the algorithms employed in FaceReader can be found in [47].

In a live analysis, FaceReader's output is a number of charts and files. Each emotion is expressed as a value between 0 and 1, indicating the intensity of the emotion. "0" means that the emotion is not visible in the facial expression, "1" means that the emotion is fully present. The emotion with the bigger value is considered the dominant one. Each time the dominant emotion changes and is active for at least 0.5 seconds, a record is written to a file. Thus, FaceReader can provide up to two different dominant emotional recordings per second. Moreover, all emotions' values (three times per second) are written at another detailed file. Furthermore, FaceReader provides all this information dynamically through live charts during its function.

In this study, only dominant emotions were taken into account. However, in our case, face modeling failed for reasons such as a student's sudden move (e.g., head rotation), a student's placing his/her hand in front of his/her face (e.g., touching his/her mouth while thinking about the answer to a question). In these cases, FaceReader did not provide any readings until the student's face or hand was back to the "correct position." Also, FaceReader provided fewer recordings of students who wore glasses or had fringes reaching down to their eyebrows. Thus, the emotional recordings we refer to are about dominant emotions that took place at time points that the aforementioned factors did not hinder the function of FaceReader.

When both the FaceReader and the reasearcher recorded a Happy, Sad, or Fear emotion, feedback was triggered manually depending on the group to which the participant was assigned. The post-feedback emotional transition was

TABLE 3
Overall, Mean and Standard Deviation Regarding Happy, Fear, and Sad Emotions

| EMOTION | APPEARENCES | | |
|---|---|---|---|
| | OVERALL | MEAN | SD |
| Happy | 258 | 1.49 | 1.84 |
| FEAR | 188 | 1.07 | 1.43 |
| SAD | 454 | 3.2 | 3.37 |

TABLE 4
Overall, Mean and Standard Deviation Regarding Feedback Messages for Happy, Fear, and Sad Emotions

| EMOTION | FEEDBACK MESSAGES | | |
|---|---|---|---|
| | OVERALL | MEAN | SD |
| Happy | 104 | 0.8 | 0.4 |
| FEAR | 85 | 0.65 | 0.47 |
| SAD | 119 | 0.91 | 0.27 |

registered only when the researcher's and the FaceReader's recordings were identical. That rule also applied to all recorded emotions.

Out of 7,416 emotional states that were recorded by the FaceReader for the 172 students, 6,440 that agreed with the researchers' observations were registered for further analysis. The interrater reliability (between each researcher and the FaceReader) for the researcher 1 and 2 was found to be $\mathrm{Kappa} = 0.72$ ($\mathrm{p} < 0.01$) and $\mathrm{Kappa} = 0.75$ ($\mathrm{p} < 0.01$), respectively. The 6,440 instances include: emotions (fear, sad, and happy) after which feedback was triggered, the affective transitions as a result of feedback, as well as emotional states during which students did not receive any feedback or emotional states that did not result from feedback. The plan for each student was to receive feedback three times during the test, once for each of the three treated emotions. This means, for instance, that if feedback was triggered for happy emotion once, it would not be triggered again, however happy the student was during the remainder of the test. Unfortunately, some students did not appear to experience all three emotions during the test. Thus, some of them received feedback only twice, or even once throughout the procedure.

Table 3 shows the overall, mean, and standard deviation regarding Happy, Fear, and Sad emotions that were observed and Table 4 presents the overall, mean and standard deviation of feedback messages for each of these three emotions.

# 5 Data Analysis

## 5.1 Metric

The computation of transition likelihoods was based on the D'Mello et al. [25] metric (1). Other studies have used this metric for affective transition analysis as well [26], [36], [48]. $L$ computes the probability of a transition between two affective states ($\mathrm{CURRENT} \rightarrow \mathrm{NEXT}$) occurring; CURRENT is a reported emotion at time $t$, and NEXT is the next reported emotion at time $t+1$:

$$L[CURRENT \rightarrow NEXT] = \frac{\frac{\Pr[NEXT \cap CURRENT]}{\Pr[CURRENT]} - \Pr[Next]}{1 - \Pr[NEXT]}. \quad (1)$$

In order to compute $L$ for each participant, the conditional probability of emotion NEXT following emotion CURRENT is calculated. In order to take into account the base rate of emotion NEXT, the D'Mello et al. metric (1) then subtracts the probability of observing emotion NEXT. $L$ accounts for the base frequency of the NEXT emotional state in assessing the likelihood of a transition. Finally, so as to normalize scores between minus infinity and 1, $L$'s numerator is divided by $1 - \Pr(NEXT)$. $L$ equal to 1 means that emotion NEXT always following the CURRENT emotion; $L$ equal to 0 means that the likelihood of emotion NEXT following the CURRENT emotion is equal to chance, i.e., the probability of observing emotion NEXT (the base rate) regardless of the CURRENT emotion. An $L$ value less than 0 means that the likelihood of emotion NEXT following the CURRENT emotion is less than chance (the probability of observing NEXT regardless of the CURRENT emotion).

## 5.2 Statistical Analysis

The aim of the following analysis was to reveal any statistically significant affective transitions from the state of a Happy, Sad, or Fear emotion to a Neutral, Sad, Happy, Disgust, Surprise, Fear, or Angry affective state that would be a result of a feedback type (N.F., ECA 1, ECA 2, ECA 3) to that emotion.

A transition's likelihood of an affective state (e.g., Happy to Neutral) as a result of one of the feedback types was examined for its significance in relation to the likelihood of this transition occurring as a result of the other feedback types. Moreover, a transition's likelihood of an affective state (e.g., Happy to Neutral) as a result of one of the feedback types was also examined for its significance in relation to the likelihood of transitioning to other affective states (e.g., Happy to Sad) as a result of that same feedback type. Thus, affective transitions were compared between the four groups, using all four levels (N.F., ECA1, ECA2, and ECA3) of factor Feedback, as well as in each group using one level of factor Feedback. The dependent variable that was measured was the likelihood of affective transitions.

An ANOVA approach would be appropriate in this case. So as to render conducting this analysis legitimate, we had to ensure that the data fulfill all the assumptions of ANOVA. First, we had to guarantee that the normality assumption was not violated. However, in this case data in all groups were far from a normal distribution. Thus, we had to use the Kruskal-Wallis test, which is the most common nonparametric equivalent of Anova. Similarly, as we would have done after an ANOVA, we used the Dunn-Sidak method (a powerful means-comparison test, similar to, but less conservative than, the Bonferroni procedure) for the post hoc comparisons. Analysis was performed using Matlab Statistical Toolbox. Thus, following a significant Kruskal-Wallis test, Dunn-Sidak multiple comparison procedure was applied to identify which transitions were significantly different, using *multcompare* function of the MATLAB Statistical Toolbox. *Multcompare's* parameter *alpha* was set for all cases at the default level (0.05), unless mentioned otherwise.

TABLE 5
Between Groups Transitions from the State of Happy

| | N.F. | ECA 1 | ECA 2 | ECA 3 |
|---|---|---|---|---|
| N | .12<br>n.s. | -.38<br>ECA 2 | -.86<br>ECA 1,<br>ECA 3 | .09<br>ECA 2 |
| Sa | .-.09 | | | |
| H | -.09<br>ECA1,<br>ECA2 | .36<br>N.F. | .67<br>N.F.,<br>ECA 3 | .18<br>ECA 2 |
| D | .03 | | | |
| Su | .005<br>n.s. | .01<br>n.s. | | -.01<br>n.s. |
| F | .01<br>n.s. | | -.03<br>n.s. | |
| A | .11<br>n.s. | -.09<br>n.s. | -.09<br>n.s. | -.03<br>n.s. |

*Between-groups analysis of transitions from the state of Happy. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Groups that statistically differ from a group at one particular transition likelihood appear in the relevant transition cell for that group.*

TABLE 6
Between Groups Transitions from the State of Sad

| | N.F. | ECA 1 | ECA 2 | ECA 3 |
|---|---|---|---|---|
| N | -.14<br>n.s. | -.48<br>n.s. | -.9<br>n.s. | -.3<br>n.s. |
| Sa | .17<br>ECA 2 | .09<br>ECA 2 | .45<br>N.F.,<br>ECA 1,<br>ECA 3 | -.05<br>ECA 2 |
| H | -.03<br>n.s. | .11<br>n.s. | .01<br>n.s. | .25<br>n.s. |
| D | .1<br>n.s. | .026<br>n.s. | | |
| Su | -.02<br>n.s. | -.04<br>n.s. | | -.027<br>n.s. |
| F | | -.02 | | |
| A | .09<br>ECA 2 | .03<br>n.s. | .011<br>N.F. | |

*Between-groups analysis of transitions from the state of Sad. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Groups that statistically differ from a group at one particular transition likelihood appear in the relevant transition cell for that group.*

## 6 Results

### 6.1 Between Groups

The factor being examined in Section 6.1 for each of the three initial emotions (Fear, Sad, and Happy) is Feedback (four levels-types of feedback) and the dependent variable being measured is the likelihood of transitioning to an emotional state as a result of being exposed to a certain level of feedback.

#### *6.1.1 Transitions from the State of Happy (Table 5)*

**Happy to neutral.** Kruskal-Wallis yields statistically significant differences (Kruskal-Wallis $X^2 = 19.3$; $p = 0.0002$) in the likelihood that a Neutral emotional state would follow a Happy emotion as a result of the four different feedback types. Dunn-Sidak multiple comparisons showed that ECA 3 has a significantly higher likelihood (0.09) to induce a Neutral emotional state after a Happy emotion than ECA 2 ($-0.86$). ECA 2 was also statistically different from ECA 1 ($-0.38$). ECA 1 was statistically indistinguishable from all other feedback types. The N.F. group was statistically indistinguishable from all feedback types.

**Happy to happy.** Statistically significant differences (Kruskal-Wallis $X^2 = 300.84$; $p = 0.000001$) were found in the likelihood that a Happy emotional state would ensue after a Happy emotion as a result of the four different feedback types. Dunn-Sidak multiple comparisons showed that the ECA 2 has a significantly higher likelihood (0.67) to make a Happy emotional state persist than the ECA 3 (0.18) and the N.F. group ($-0.09$). ECA 3 only statistically differed from the ECA 2. ECA 1 (0.36) only statistically differed from the N.F. group ($-0.09$).

**Happy to sad-disgust-surprise-fear-anger.** A Sad or a Disgust emotional state was not observed when ECA 1, ECA 2, and ECA 3 were displayed after a Happy emotional state. Also, a Fear emotional state was not observed when ECA 1 and ECA 3 were displayed after a Happy emotion. Moreover, a Surprise emotional state was not observed when ECA 2 was displayed after a Happy emotional state. Besides, Kruskal-Wallis yields no statistically significant differences in the likelihood that a Surprise, Fear, or Anger emotional state would follow a Happy emotional state as a result of the different feedback types.

#### *6.1.2 Transitions from the State of Sad (Table 6)*

**Sad to sad.** Kruskal-Wallis yields statistically significant differences (Kruskal-Wallis $X^2 = 320.42$; $p = 0.000001$) in the likelihood that a Sad emotional state would follow a Sad emotion as a result of the four different feedback groups. Dunn-Sidak multiple comparisons showed that the ECA 2 has a significantly higher likelihood (0.45) to make a Sad emotional state persist than the N.F. group (0.17), ECA 1 (0.09), and ECA 3 ($-0.05$). The N.F. group, ECA 1, and ECA 3, statistically differed only toward ECA 2.

**Sad to angry.** Statistically significant differences (Kruskal-Wallis $X^2 = 7.54$; $p = 0.02$) were found in the likelihood that an Angry emotional state would follow a Sad emotion as a result of the three different feedback types (N.F., ECA 1, and ECA 2). Anger was not observed when ECA 3 was displayed after a Sad emotional state. Dunn-Sidak multiple comparisons confirmed that the N.F. group has a significantly higher likelihood (0.09) to induce an Angry emotional state after a Sad emotion than ECA 2 (0.011). ECA 1 was statistically indistinguishable.

**Sad to neutral-happy-disgust-surprise-fear.** A Disgust, Surprise, or Fear emotional state was not observed when ECA 2 was displayed after a Sad emotional state. Moreover, a Fear emotional state after a Sad emotion was not observed in the N.F. group. Besides, Kruskal-Wallis yields no statistically significant differences in the likelihood that a Neutral, Happy, Disgust, or Surprise emotional state would follow a Sad emotional state as a result of the different feedback types.

TABLE 7
Between Groups Transitions from the State of Fear

| | N.F. | ECA 1 | ECA 2 | ECA 3 |
|---|---|---|---|---|
| N | -.32 ECA 3 | -.65 ECA 3 | -.98 ECA 3 | .83 N.F., ECA 1, ECA 2 |
| Sa | | -.07 | | |
| H | -.03 n.s. | .14 n.s | .04 n.s. | .06 n.s. |
| D | .05 n.s. | .005 n.s | | .01 n.s. |
| Su | .03 n.s. | -.001 n.s | | |
| F | .14 ECA 2 | .012 ECA 2 | .57 N.F., ECA 1 | |
| A | .09 n.s | .16 n.s | -.08 n.s. | |

*Between-groups analysis of transitions from the state of Fear. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Groups that statistically differ from a group at one particular transition likelihood appear in the relevant transition cell for that group.*

### 6.1.3 *Transitions from the State of Fear (Table 7)*

**Fear to neutral.** Kruskal-Wallis yields statistically significant differences (Kruskal-Wallis $X^2 = 20.68$; $p = 0.0001$) in the likelihood that a Neutral emotional state would follow a Fear emotion as a result of the four different feedback types. Dunn-Sidak multiple comparisons showed that ECA 3 has a significantly higher likelihood (0.83) to induce a Neutral emotional state when displayed after a Fear emotion than the N.F. group ($-0.32$), ECA 1 ($-0.65$), and ECA 2 ($-0.98$). The N.F. group, ECA 1, and ECA 2, statistically differed only toward ECA 3.

**Fear to fear.** Statistically significant differences (Kruskal-Wallis $X^2 = 28.1$; $p = 0.000001$) were found in the likelihood that a Fear emotional state would follow a Fear emotion as a result of the three different feedback groups (N.F., ECA 1, and ECA 2). Fear was not observed when ECA 3 was displayed after a Fear emotional state. Dunn-Sidak multiple comparisons showed that ECA 2 exhibits a significantly higher likelihood (0.57) of making a Fear emotional state persist than the N.F. group (0.14), and ECA 1 (0.012). The N.F. group and the ECA 1 statistically differed only from ECA 2.

**Fear to sad-happy-disgust-surprise-angry.** A Sad emotional state was not observed when the ECA 2 or ECA 3 was displayed after an emotional state of Fear. Moreover, a Sad emotional state after a Fear emotion was not observed in the N.F. group. A Disgust emotional state was not observed when ECA 2 was displayed after a Fear emotion. A Surprise emotional state was not observed when ECA 2 or ECA 3 was displayed after an emotional state of Fear. An Angry emotional state was not observed when ECA 3 was displayed after an emotional state of Fear. Besides, Kruskal-Wallis yields no statistically significant differences in the likelihood that a Happy, Disgust, Surprise, or Angry emotional state would follow a Fear emotional state, as a result of the different feedback types.

TABLE 8
Transitions in Each Group from the State of Happy

| | N | Sa | H | D | Su | F | A |
|---|---|---|---|---|---|---|---|
| N.F. | .12 n.s. | .-.09 n.s. | -.09 n.s. | .03 n.s. | .005 n.s. | .01 n.s. | .11 n.s. |
| ECA1 | -.38 A | | .36 A | | .01 n.s. | | -.09 H, N |
| ECA2 | -.86 F,H | | .67 A,F,N | | | -.03 H,N | -.09 H |
| ECA3 | .09 H,Su,A | | .18 N | | -.01 N | | -.03 N |

*Transitions in each group from the state of Happy. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Transitions in each group that statistically differ from one transition likelihood appear in the relevant transition cell for that group (e.g., H, Su, A).*

## 6.2 In Each Group

In addition to the analysis performed in Section 6.1, in order to test the null hypothesis that the difference in the likelihoods of transitions (e.g., Happy to Neutral versus Happy to Sad) was not statistically significant in each group, a one-level Kruskal-Wallis (nonparametric anova) was performed for every group. The factor being examined for each of the three initial emotions (Fear, Sad, and Happy) is again Feedback, but now seen only at one level, using transitions to the different emotional states as subgroups.

### 6.2.1 *Transitions from the State of Happy (Table 8)*

**N.F. group.** Transitions from the state of Happy were not significantly different in the N.F. group.

**ECA 1 group**. Analyzing the transitions from the state of Happy we find that ECA 1 yields statistically significant differences (Kruskal-Wallis $X^2 = 13.96$; $p = 0.003$) in the likelihood that the Neutral, Happy, Surprise, and Angry emotions may occur as a result of this emotional feedback type. Transitions to Sad, Disgust, and Fear emotions did not occur. Dunn-Sidak multiple comparisons showed that Happy (0.36) has a significantly higher likelihood of following Happy than does Angry ($-0.09$). However, the transition to a Happy emotion was statistically indistinguishable from that of Surprise (0.01) and Neutral ($-0.38$). Moreover, no transition was statistically different from that of Surprise. Transition to Neutral statistically differed from that of Angry and transition to Angry statistically differed from that of Neutral and Happy.

**ECA 2 group.** Analyzing the transitions from the state of Happy we find that ECA 2 yields statistically significant differences (Kruskal-Wallis $X^2 = 68.84$; $p = 0.000000000001$) in the likelihood that Neutral, Happy, Fear, and Angry may occur as a result of this emotional feedback type. Sad, Disgust, and Surprise transitions did not occur. Dunn-Sidak multiple comparisons showed that Happy has a significantly higher likelihood (0.67) of following Happy than does Fear ($-0.03$), Angry ($-0.09$), and Neutral ($-0.86$). Moreover, the transition to Angry statistically differed only from that of Happy, Neutral differed from Fear and Happy, and Fear differed from Neutral and Happy.

**ECA 3 group.** Analyzing the transitions from the state of Happy we find that ECA 3 yields statistically significant differences (Kruskal-Wallis $X^2 = 18.11$; $p = 0.0004$) in the

TABLE 9
Transitions in Each Group from the State of Sad

| | N | Sa | H | D | Su | F | A |
|---|---|---|---|---|---|---|---|
| N.F. | -.14 n.s. | .17 A,D,Su,H | -.03 Sa | .1 Sa | -.02 Sa | | .09 Sa |
| ECA1 | -.48 n.s. | .09 n.s | .11 n.s. | .026 n.s. | -.04 n.s. | -.02 n.s. | .03 n.s. |
| ECA2 | -.9 Sa | .45 A,N,H | .01 Sa | | | | .011 Sa |
| ECA3 | -.3 n.s. | -.05 H,Su | .25 Sa | | -.027 Sa | | |

*Within-groups analysis of transitions from the state of Sad. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Transitions in each group that statistically differ from one transition likelihood appear in the relevant transition cell for that group (e.g., H, Su, A).*

TABLE 10
Transitions in Each Group from the State of Fear

| | N | Sa | H | D | Su | F | A |
|---|---|---|---|---|---|---|---|
| N.F. | -.32 n.s. | | -.03 F. | .05 F. | .03 F. | .14 D.,Su,H | .09 n.s. |
| ECA1 | -.65 n.s. | -.07 n.s. | .14 n.s. | .005 n.s. | -.001 n.s. | .012 n.s. | .16 n.s. |
| ECA2 | -.98 H,F | | .04 N,F | | | .57 N,H,A | -.08 F |
| ECA3 | .83 H,D | | .06 N | .01 N | | | |

*Within-groups analysis of transitions from the state of Fear. N = Neutral, Sa = Sad, H = Happy, D = Disgust, Su = Surprise, F = Fear, A = Angry, n.s = Not Significant, empty spaces = nothing observed. Numbers in cells denote the likelihood of a transition occurring. Note: Transitions in each group that statistically differ from one transition likelihood appear in the relevant transition cell for that group (e.g., H, Su, A).*

likelihood that the Neutral, Happy, Surprise, and Angry emotions may occur after a Happy emotion as a result of this emotional feedback type. Transitions to Sad, Disgust, and Fear emotions were not observed when ECA 3 was displayed after a Happy emotion. For this feedback type, Dunn-Sidak multiple comparisons indicated that Happy exhibits a significantly higher likelihood (0.18) of following Happy than does Neutral (0.09), while Surprise (−0.01) and Angry (−0.03) transitions did not significantly differ from that of Happy. However, Neutral was statistically different from Happy, Surprise, and Angry. Surprise and Angry only differed from Neutral.

### 6.2.2 Transitions from the State of Sad (Table 9)

**N.F. group.** Kruskal-Wallis test showed statistically significant differences (Kruskal-Wallis $X^2 = 29.02$; $p = 0.0001$) in the N.F. group in the likelihood that Neutral, Sad, Happy, Disgust, Surprise, and Anger would follow a Sad emotion. Fear was not observed after Sad at the N.F. group. Dunn-Sidak multiple comparisons showed that Sad has a significantly higher likelihood (0.17) of following Sad than does Angry (0.09), Disgust (0.1), Surprise (−0.02), and Happy (−0.03). The transition to Neutral (−0.14) was statistically indistinguishable from all transitions. Transitions to Angry, Disgust, Surprise, and Happy significantly differed only from the transition to Sad.

**ECA 1 group.** Kruskal-Wallis test showed that ECA 1 yields no statistically significant differences in the likelihood that emotional transitions would follow a Sad emotion as a result of this emotional feedback type.

**ECA 2 group.** Kruskal-Wallis test showed that ECA 2 yields statistically significant differences ($X^2 = 23.88$; $p = 0.0001$) in the likelihood that Neutral, Sad, Happy, and Anger would follow a Sad emotion as a result of this emotional feedback type. Disgust, Surprise, and Fear emotions were not observed when ECA 2 was displayed after a Sad emotion. Dunn-Sidak multiple comparisons showed that Sad has a significantly higher likelihood (0.45) of following Sad than Angry (0.011) and Neutral (−0.9). Happy transition likelihood (0.01) would only be statistically different from that to Sad at a 90 percent confidence level. Angry, Neutral, and Happy transitions differed significantly only from the transition to Sad.

**ECA 3 group.** Kruskal-Wallis test showed that ECA 3 yields statistically significant differences ($X^2 = 13.86$; $p = 0.0031$) in the likelihood that Neutral, Sad, Happy, and Surprise would follow a Sad emotion as a result of this emotional feedback type. Fear, Disgust, and Angry emotions were not observed when ECA 3 was displayed as feedback to Sad. Dunn-Sidak multiple comparisons showed that Happy has a significantly higher likelihood (0.25) of following Sad than does Sad (−0.05). Transition to Neutral (−0.3) was statistically indistinguishable from all transitions. Transition to Surprise (−0.027) only differed significantly from that to Sad.

### 6.2.3 Transitions from the State of Fear (Table 10)

**N.F. group.** Transitions from the state of Fear at the nonfeedback group were significantly different (Kruskal-Wallis $X^2 = 17.47$; $p = 0.037$) in the likelihood that Neutral, Happy, Disgust, Surprise, Fear, and Anger would follow Fear. Sad was not observed following Fear in the N.F. group. Dunn-Sidak multiple comparison set at 90 percent confidence level showed that Fear has a significantly higher likelihood (0.14) of following Fear than does Disgust (0.05), Surprise (0.03), and Happy (−0.03). The transition to Fear was statistically indistinguishable from that of Angry (0.09) and Neutral (−0.31), though Angry and Neutral transitions were not significantly different from any transition. Disgust, Surprise, and Happy transitions differed significantly only from that to Fear.

**ECA 1 group.** Transitions from the state of Fear were not significantly different in the likelihood that emotional transitions would follow Fear as a result of ECA 1.

**ECA 2 group.** Transitions from the state of Fear are significantly different (Kruskal-Wallis $X^2 = 40.9$; $p = 0.00000001$) in the likelihood that Neutral, Happy, Fear, and Angry emotions would follow Fear as result of ECA 2. Transitions to Sad, Disgust, and Surprise emotions were not observed when ECA 2 was displayed as feedback to Fear. Dunn-Sidak multiple comparisons showed that Fear has a significantly higher likelihood (0.57) of following Fear than does Happy (0.04), Angry (−0.08), and Neutral (−0.98). Happy was significantly different from Neutral and Fear, Angry only differed significantly from Fear, while Neutral was significantly different from Happy and Fear.

**ECA 3 group.** Transitions from the state of Fear are significantly different (Kruskal-Wallis $X^2 = 26.27$; $p = 0.00001$) in the likelihood that Neutral, Happy, and Disgust emotions would follow Fear as result of ECA 3. Transitions to Sad, Surprise, Fear, and Angry were not observed when ECA 3 was displayed as feedback to Fear. Dunn-Sidak multiple comparisons showed that Neutral has a significantly higher likelihood (0.83) of following Fear than does Happy (0.06) and Disgust (0.01). Transitions to Happy and Disgust were only significantly different from the transition to Neutral.

## 7 Conclusion

An important finding of this study is that the ECA 2 (performing parallel empathy displaying an emotional expression that was relevant to the emotional state of the student) reinforced the student's emotion.

When ECA 2 was displayed while the student was experiencing Fear, there was a great likelihood (0.57) that the student would remain in that state. This result was statistically different from the relevant result for the ECA 1 (0.012) and the N.F. (0.14) group (Fear was not observed when ECA 3 was displayed as feedback to Fear). Moreover, the likelihood of this transition (Fear to Fear) as a result of the ECA 2 also statistically differed from the likelihood of transitioning to the state of Neutral ($-0.98$), Happy (0.04), and Angry ($-0.08$) as a result of displaying ECA 2 as feedback to the emotion of Fear (transitions to Sad, Disgust, and Surprise were not observed in this case).

Similarly, when ECA 2 was displayed while the student was Sad, there was a great likelihood (0.45) that the student would remain Sad. This result was statistically different from the relevant result for the ECA 1 (0.09), the ECA 3 ($-0.05$), and the N.F. (0.17) group. Furthermore, this transition's likelihood (Sad to Sad), as a result of the ECA 2, also statistically differed from the likelihood to transitioning to the state of Neutral ($-0.9$), Happy (0.01), and Angry (0.011) as a result of displaying ECA 2 as feedback to a Sad student (transitions to Disgust, Surprise, and Fear were not observed in this case).

When ECA 2 was displayed while the student was Happy, there was a great likelihood (0.67) that the student would remain Happy. This result was statistically different from the relevant result for the ECA 3 (0.18) and the N.F. ($-0.09$) group, but it was statistically indistinguishable from the ECA 1 (0.36). In addition, this transition's likelihood (Happy to Happy), as a result of the ECA 2, also statistically differed from the likelihood to transitioning to the state of Neutral ($-0.86$), Fear ($-0.03$), and Angry ($-0.09$) as a result of displaying ECA 2 as feedback to a Happy student (transitions to Sad, Disgust, and Surprise were not observed in this case).

Individuals react as in a social context to both human and computer-controlled entities [5], [6]. Moreover, there is research evidence suggesting that users can experience empathetic emotional reactions toward embodied agents [49]. In the ECA 2 case, students showed empathy toward the agent's emotion by expressing that same emotion themselves. Nevertheless, this was not observed for the ECA 1, having the same appearance and performing the same empathetic behavior as the ECA 2, but displaying a neutral emotional expression. Consequently, it must be the emotional expression that made the difference between ECA 1 and ECA 2.

Strong transitions to the same emotional state were observed when ECA 2 was displayed as feedback to Fear, Sad, and Happy emotions. Possibly that effect could be observed with other emotional states as well. For instance, in [37] it is stated that students experiencing frustration are very likely to remain frustrated if presented with a character mimicking their emotions by displaying parallel empathetic behavior. Moreover, an agent's parallel empathetic behavior has been shown [50] to reinforce a student's experience of boredom. Nevertheless, it has also been shown in [50] that an agent displaying parallel empathetic behavior may encourage students experiencing the state of flow to remain in that state, and thus remain in that "virtuous cycle" of learning. Thus, future developers may need to take into account this effect when designing empathetic agents for tutoring systems.

Another important finding of this study is that the ECA 3 (performing a combination of parallel and reactive empathy, displaying an emotional expression that was relevant to the emotional state of the student for parallel empathy, and subsequently displaying an emotional expression distinct from the emotional state of the student for reactive empathy) appeared to be considerably likely (0.83) to induce a student to Neutral state when displayed as feedback to a student experiencing Fear. This result was statistically different from the relevant result for the ECA 1 ($-0.65$), the ECA 2 ($-0.98$), and the N.F. ($-0.32$) group. Moreover, the likelihood of this transition (Fear to Neutral), as a result of the ECA 3, also statistically differed from the likelihood of transitioning to the state of Happy (0.06) and Disgust (0.01) as a result of displaying ECA 3 as feedback to the emotion of Fear (transitions to Sad, Surprise, Fear, and Angry were not observed in this case).

ECA 3 appeared to be effective in altering the state of Fear to a Neutral one. A similar effect was not observed when ECA 3 was displayed as feedback to Sad and Happy emotions. It is unclear whether this has to do with the particular nature of Sad and Happy emotions or with other factors as well.

This is the first time that agents' empathetic behavior has been examined in combination with agents' emotional facial and tone of voice expressions within a learning context. Thus, what is essential here is the evidence that this kind of reactive behavior (ECA 3) does have an effect on students' emotional regulation, even if this effect proved to be statistically significant only for Fear. The ECA 3 could also have been effective in regulating the emotional states of sad and happy emotions if its verbal behavior and emotional expressions were different. Besides, it has been shown in [50] that agents displaying reactive empathetic behavior are most likely to alter students' emotional states toward learning goals. For the time being, we can retain the ECA 3 effect on Fear as an empirical finding. Since empathy is as yet only partially comprehended, it is unclear which kind of empathetic behavior and under what conditions would be most effective. Thus, empirical findings from human-ECA interaction could contribute to developing agents that would respond appropriately in various social contexts [36].

Other findings of this study were not deemed considerable because of their limited likelihood and/or their lack of statistical significance. However, all findings are presented in Section 6. On some occasions even insignificant findings can provide useful insight.

In the future, we plan to experiment with emotional states other than basic emotions to meet the needs of personalized self-assessment. Affective feedback should be able to adapt even when learners experience uncommon emotional states. Furthermore, we plan to examine ECAs' empathetic behavior in combination with emotional facial expressions and body movements. A valuable platform for doing this kind of work is GRETA, which has been used in various European projects (CALLAS, SEMAINE, HUMAINE). GRETA is an embodied conversational agent with a 3D representation of a woman, accustomed to MPEG-4 animation standard. "She" is capable of talking and at the same time performing facial expressions, gestures, gaze, and head movements [51]. The platform provides the researcher with numerous options of verbal and nonverbal behaviors, plus the opportunity to implement his/her own new patterns of ECA behavior.

Research toward affective learning systems is an extremely important multidisciplinary area, involving a collaborative effort on the part of scientists from various fields. Students with great talent in science or other disciplines could fail to perceive their potential and follow a career that does not reflect their real inclination because of emotional exhaustion, lack of self-competence, or other psychological inefficiencies. Consequently, an educational system not integrating helpful affective learning techniques could result in unsatisfied individuals and in loss of valuable social capital.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the editor Jonathan Gratch and three anonymous reviewers for their valuable comments and suggestions.

## REFERENCES

[1] J.N. Bailenson, N. Yee, D. Merget, and R. Schroeder, "The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction," *Presence: Teleoperators and Virtual Environments*, vol. 15, pp. 359-372, 2006.
[2] J. Cassell and P. Miller, "Is It Self-Administration If the Computer Gives You Encouraging Looks?" *Envisioning the Survey Interview of the Future*, F.G. Conrad and M.F. Schober, eds., pp. 161-178, John Wiley & Sons, 2007.
[3] C.N. Moridis and A.A. Economides, "Towards Computer-Aided Affective Learning Systems: A Literature Review," *J. Educational Computing Research*, vol. 39, no. 4, pp. 313-337, 2008.
[4] C.R. Rogers, "A Theory of Therapy, Personality and Interpersonal Relationships, as Developed in the Client-Centered Framework," *Psychology: A Study of Science*, S. Koch, ed. McGraw-Hill, 1959.
[5] C. Nass and Y. Moon, "Machines and Mindlessness: Social Responses to Computers," *J. Social Issues*, vol. 56, no. 1, pp. 81-103, 2000.
[6] B. Reeves and Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge Univ. Press, 1996.
[7] S. Brave, C. Nass, and K. Hutchinson, "Computers that Care: Investigating the Effects of Orientation of Emotion Exhibited by an Embodied Computer Agent," *Int'l J. Human-Computer Studies*, vol. 62, no. 2, pp. 161-178, 2005.
[8] D.M. Dehn and S. Van Mulder, "The Impact of Animated Interface Agents: A Review of Empirical Research," *Int'l J. Human-Computer Studies*, vol. 52, no. 1, pp. 1-22, 2000.
[9] M.H. Davis, *Empathy: A Social Psychological Approach.* Westview Press, 1996.
[10] M.Z. Yusoff and B. Du Boulay, "The Integration of Domain-Independent Strategies into an Affective Tutoring System: Can Students' Learning Gain Be Improved?" *Electronic J. Computer Science & Information Technology*, vol. 1, no. 1, 2009.
[11] C. Achebe, "Multi-Modal Counselling for Examination Failure in a Nigerian University: A Case Study," *J. African Studies*, vol. 9, pp. 187-193, 1982.
[12] A. Efklides and S. Volet, "Feelings and Emotions in the Learning Process," *Learning and Instruction*, vol. 15, no. 5, pp. 1-10, 2005.
[13] A.A. Economides, "Personalized Feedback in CAT (Computer Adaptive Testing)," *Trans. Advances in Eng. Education*, vol. 2, no. 3, pp. 174-181, 2005.
[14] S. Oviatt, "User-Modeling and Evaluation of Multimodal Interfaces," *Proc. IEEE*, special issue on human-computer multimodal interface, vol. 91, no. 9, pp. 1457-1468, Sept. 2003.
[15] M. Pantic and L.J.M. Rothkrantz, "Toward an Affect-Sensitive Multimodal Human-Computer Interaction," *Proc. IEEE*, special issue on human-computer multimodal interface, vol. 91, no. 9, pp. 1370-1390, Sept. 2003.
[16] M.J. Den Uyl and H. van Kuilenburg, "The FaceReader: Online Facial Expression Recognition," *Proc. Measuring Behaviour*, pp. 589-590, 2005.
[17] P. Ekman and W.V. Friesen, *Manual for the Facial Action Coding System.* Consulting Psychologists Press, 1977.
[18] C. Conati, "Probabilistic Assessment of User's Emotions during the Interaction with Educational Games," *J. Applied Artificial Intelligence*, special issue on merging cognition and affect in HCI, vol. 16, pp. 555-575, 2002.
[19] C. Conati and H. Maclaren, "Empirically Building and Evaluating a Probabilistic Model of User Affect," *User Modeling and User-Adapted Interaction*, vol. 19, pp. 267-303, 2009.
[20] A. Ortony, G.L. Clore, and A. Collins, *The Cognitive Structure of Emotions.* Cambridge Univ. Press, 1988.
[21] R.A. Calvo and S.D. Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods and Their Applications," *IEEE Trans. Affective Computing*, vol. 1, no. 1, pp. 18-37, Jan.-June 2010.
[22] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and Learning: An Exploratory Look into the Role of Affect in Learning with AutoTutor," *Learning, Media and Technology*, vol. 29, pp. 241-250, 2004.
[23] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard, "Affect-Aware Tutors: Recognising and Responding to Student Affect," *Int'l J. Learning Technology*, vol. 4, no. 3/4, pp. 129-164, 2009.
[24] R. Baker, S. D'Mello, M. Rodrigo, and A. Graesser, "Better to Be Frustrated than Bored: The Incidence and Persistence of Affect during Interactions with Three Different Computer-Based Learning Environments," *Int'l J. Human-Computer Studies*, vol. 68, no. 4, pp. 223-241, 2010.
[25] S. D'Mello, R. Taylor, and A. Graesser, "Monitoring Affective Trajectories during Complex Learning," *Proc. 29th Ann. Cognitive Science Soc.*, pp. 203-208, 2007.
[26] S. D'Mello, T. Jackson, S. Craig, B. Morgan, P. Chipman, H. White, N. Person, B. Kort, R. el Kaliouby, R. Picard, and A. Graesser, "AutoTutor Detects and Responds to Learners Affective and Cognitive States," *Proc. Workshop Emotional and Cognitive Issues in ITS in Conjunction with the Ninth Int'l Conf. Intelligent Tutoring Systems*, pp. 31-43, 2008.
[27] S. McQuiggan, J. Robison, and J. Lester, "Affective Transitions in Narrative-Centered Learning Environments," *Proc. Ninth Int'l Conf. Intelligent Tutoring Systems*, 2008.
[28] Y. Kim, "Empathetic Virtual Peers Enhanced Learner Interest and Self-Efficacy," *Proc. Workshop Motivation and Affect in Educational Software at the 12th Int'l Conf. Artificial Intelligence in Education*, pp. 9-16, 2005.
[29] J.C. Lester, S.G. Towns, and P.J. FitzGerald, "Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents," *Int'l J. Artificial Intelligence in Education*, vol. 10, nos. 3/4, pp. 278-291, 1999.
[30] K. Hone, L. Axelrod, and B. Pakekh, "Development and Evaluation of an Empathic Tutoring Agent," *Proc. Joint Symp. Virtual Social Agent: Social Intelligence and Interaction in Animals, Robots and Agents*, pp. 103-108, 2005.
[31] W. Burleson and R. Picard, "Gender-Specific Approaches to Developing Emotionally Intelligent Learning Companions," *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 62-69, July/Aug. 2007.

[32] A.C. Graesser, P. Chipman, B.C. Haynes, and A. Olney, "AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue," *IEEE Trans. Education,* vol. 48, no. 4, pp. 612-618, Nov. 2005.
[33] A.C. Graesser, N. Person, D. Harter and the Tutoring Research Group "Teaching Tactics and Dialog in AutoTutor," *Int. J. Artificial Intelligence in Education,* vol. 12, pp. 257-279, 2001.
[34] A.C. Graesser, K. VanLehn, C. Rose, P. Jordan, and D. Harter, "Intelligent Tutoring Systems with Conversational Dialogue," *AI Magazine,* vol. 22, pp. 39-51, 2001.
[35] A.C. Graesser, K. Wiemer-Hastings, P. Wiemer-Hastings, R. Kreuz, andthe Tutoring Research Group, "AutoTutor: A Simulation of a Human Tutor," *J. Cognitive Systems Research,* vol. 1, pp. 35-51, 1999.
[36] S. McQuiggan, J. Robison, R. Phillips, and J. Lester, "Modeling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach," *Proc. Seventh Int'l Joint Conf. Autonomous Agents and Multi-Agent Systems,* 2008.
[37] J. Robison, S. McQuiggan, and J. Lester, "Differential Affective Experiences in Narrative-Centered Learning Environments," *Proc. Workshop Emotional and Cognitive issues in ITS in Conjunction with the Ninth Int'l Conf. Intelligent Tutoring Systems,* 2008.
[38] N. Yee, J. Bailenson, and K. Rickertsen, "A Meta-Analysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces," *Proc. SIGCHI Conf. Human Factors in Computing Systems,* M.B. Rosson and D. Gilmore, eds., pp. 1-10, 2007.
[39] K. Hone, "Empathic Agents to Reduce User Frustration: The Effects of Varying Agent Characteristics," *Interacting with Computers,* vol. 18, no. 2, pp. 227-245, 2005.
[40] E.J. Lee, C. Nass, and S. Brave, "Can Computer-Generated Speech Have Gender? An Experimental Test of Gender Stereotype," *Proc. Conf. Human Factors in Computing Systems Extended Abstracts on Human Factors in Computing Systems,* pp. 289-290, 2000.
[41] H. Prendinger and M. Ishizuka, "The Empathic Companion: A Character-Based Interface that Addresses Users' Affective States," *Int'l J. Applied Artificial Intelligence,* vol. 19, pp. 285-297, 2005.
[42] M. Ochs, C. Pelachaud, and D. Sadek, "An Empathic Virtual Dialog Agent to Improve Human-Machine Interaction," *Proc. Seventh Int'l Joint Conf. Autonomous Agents and Multi-Agent Systems,* 2008.
[43] R. Niewiadomski, M. Ochs, and C. Pelachaud, "Expressions of Empathy in eCAs," *Proc. 10th Int'l Conf. Intelligent Virtual Agents,* H. Prendinger, J.C. Lester, and M. Ishizuka, eds., pp. 37-44, 2008.
[44] C.N. Moridis and A.A. Economides, "Prediction of Student's Mood during an Online Test Using Formula-Based and Neural Network-Based Method," *Computers & Education,* vol. 53, no. 3, pp. 644-652, 2009.
[45] C.N. Moridis and A.A. Economides, "Mood Recognition during Online Self-Assessment Tests," *IEEE Trans. Learning Technologies,* vol. 2, no. 1, pp. 50-61, Jan./Mar. 2009.
[46] B. Zaman and T. Shrimpton-Smith, "The FaceReader: Measuring Instant Fun of Use," *Proc. Fourth Nordic Conf. Human-Computer Interaction,* pp. 457-460, 2006.
[47] H. van Kuilenburg, M. Wiering, and M. den Uyl, "A Model Based-Method for Automatic Facial Expression Recognition," *Proc. 16th European Conf. Machine Learning,* 2005.
[48] R. Baker, M. Rodrigo, and U. Xolocotzin, "The Dynamics of Affective Transitions in Simulation Problem-Solving Environments," *Proc. Second Int'l Conf. Affective Computing and Intelligent Interactions,* pp. 666-677, 2007.
[49] A. Paiva, J. Dias, D. Sobral, S. Woods, and L. Hall, "Building Empathic Lifelike Characters: The Proximity Factor," *Proc. Int'l Joint Conf. Autonomous Agents and Multi-Agent Systems,* 2004.
[50] S.W. McQuiggan, J.L. Robison, and J.C. Lester, "Affective Transitions in Narrative-Centered Learning Environments," *Educational Technology and Soc.,* vol. 13, no. 1, pp. 40-53, 2010.
[51] C. Pelachaud, "Studies on Gesture Expressivity for a Virtual Agent," *Speech Comm.,* vol. 51, pp. 630-639, 2009.

**Christos N. Moridis** received the BS degree in communication, media, and culture from the Panteion University of Athens in 2004, the MSc degree in advanced systems of computing and communications, specializing in intelligent systems, from the Aristotle University of Thessaloniki in 2007, and the PhD degree in information systems from the University of Macedonia, Thessaloniki, Greece, in 2011. His research interests include affective learning systems, emotional agents, neural networks and fuzzy systems for emotion modeling, interface design, and emotional assessment using electroencephalography (EEG). He is a member of the IEEE, the IEEE Computer Society, and the HUMAINE Association.

**Anastasios A. Economides** received the DiplEng degree in electrical engineering from Aristotle University of Thessaloniki. Holding a Fulbright and a Greek State Fellowship, he received the MSc and the PhD degrees in computer engineering from the University of Southern California, Los Angeles. He is a professor of computer networks and telematic applications and the chairman of the Information Systems Postgraduate Program at the University of Macedonia, Thessaloniki, Greece. His research interests include networking techno-economics, e-learning, e-tourism, and e-government. He has published more than 170 peer-reviewed papers and has received more than 400 citations. He is a senior member of the IEEE.