

The Relationship between Affective States and  
Dialog Patterns during Interactions with AutoTutor

Arthur C. Graesser<sup>1</sup>, Sidney K. D'Mello<sup>1</sup>, Scotty D. Craig<sup>2</sup>, Amy Witherspoon<sup>1</sup>,  
Jeremiah Sullins<sup>1</sup>, Bethany McDaniel<sup>1</sup>, & Barry Gholson<sup>1</sup>

<sup>1</sup> University of Memphis

<sup>2</sup> University of Pittsburgh

Send correspondence to:

Arthur C. Graesser  
Department of Psychology and Institute for Intelligent Systems  
202 Psychology Building  
University of Memphis  
Memphis, TN 38152-3230  
(901) 678-2146  
(901) 678-2579 (fax)  
[a-graesser@memphis.edu](mailto:a-graesser@memphis.edu)

running head: AUTOTUTOR AND EMOTIONS

### Abstract

Relations between emotions (affect states) and learning have recently been explored in the context of AutoTutor. AutoTutor is a tutoring system on the Internet that helps learners construct answers to difficult questions by interacting with them in natural language. AutoTutor has an animated conversation agent and a dialog management facility that attempts to comprehend the learner's contributions and to respond with appropriate dialog moves (such as short feedback, pumps, hints, prompts for information, assertions, answers to student questions, suggestions for actions, summaries). Our long-term goal is to build an adaptive AutoTutor that responds to the learners' affect states in addition to their cognitive states. The present study adopted an *emote-aloud* procedure in which participants were videotaped as they verbalized their affective states (called *emotes*) while interacting with AutoTutor on the subject matter of computer literacy. The emote-aloud protocols uncovered a number of affective states (notably confusion, frustration, and eureka/delight). The AutoTutor log files were mined to identify characteristics of the dialogue and learners' knowledge states that were correlated with these affect states. We report the significant correlations and speculate on their implications for the larger project of building a non-intrusive, affect-sensitive AutoTutor.

Connections between emotions and complex learning are receiving more attention in the fields of psychology (Carver, 2004; Deci & Ryan, 2002; Dweck, 2002), education (Lepper & Henderlong, 2000; Linnenbrink & Pintrich, 2002; Meyer & Turner, 2002), neuroscience (Damasio, 2003), and computer science (Kort, Reilly, & Picard, 2001; Picard, 1997). A satisfactory understanding of such emotion-learning connections is needed to design engaging educational artifacts. Such artifacts include adaptive intelligent tutoring systems on technical material (DeVicente & Pain, 2002; Graesser, Person, Lu, Jeon, & McDaniel, 2005; Guhe, Gray, Schoelles, & Ji, 2004; Litman & Silliman, 2004), serious games (Conati, 2002; Gee, 2003), and non-interactive media (Vorderer, 2003).

Psychologists have developed theories that link cognition and emotions very generally (Bower, 1981; Mandler, 1984; Ortony, Clore, & Collins, 1988; Russell, 2003; Stein & Levine, 1991). These theories convey general links between cognition and emotions (affect), but they do not directly explain and predict the emotions that occur during complex learning, such as attempts to master physics, biology, computer literacy, or critical thinking. Researchers are familiar with Ekman's work on the detection of emotions from facial expressions (Ekman, 2003; Ekman & Friesen, 1978). However, the emotions that Ekman intensely investigated (e.g., sadness, happiness, anger, fear, disgust, surprise) have minimal relevance to learning per se (Graesser, McDaniel, Chipman, Witherspoon, D'Mello, 2006; Fredrickson & Branigan, 2005; Kort et al., 2001; Schutzwohl & Borgstedt, 2005). The pervasive affective states during complex learning include confusion, boredom, flow/engagement, curiosity/interest, delight/eureka, and frustration from being stuck (Burlison & Picard, 2004; Craig, Graesser, Sullins, & Gholson, 2004; Csikszentmihalyi, 1990; Graesser et al., 2006; Kort et al., 2001).

There are a number of ways in which tutors and other types of learning environments might adaptively respond to the learner's emotions during the course of enhancing learning (D'Mello, Craig, Gholson, Franklin, Picard, & Graesser, 2005; Graesser, Jackson, & McDaniel, in press; Lepper & Woolverton, 2002). If the learner is frustrated, for example, the tutor can give hints to advance the learner in constructing knowledge or can make supportive empathetic comments to enhance motivation (Burlison & Picard, 2004). If the learner is bored, the tutor needs to present more engaging or challenging problems for the learner to work on. The tutor would probably want to lay low and stay out of the learner's way when the learner is in a state of flow (Csikszentmihaly, 1990), i.e., when the learner is so deeply engaged in learning the material that time and fatigue disappear. The flow experience is believed to occur when the learning rate is high and the learner has achieved a high level of mastery at the *region of proximal learning* (Metcalfe & Kornell, 2005).

The affective state of confusion is particularly interesting because it is believed to play an important role in learning (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Guhe et al., 2004) and has a significant positive correlation with learning gains (Craig, Graesser et al., 2004). Confusion is diagnostic of *cognitive disequilibrium*, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Festinger, 1957; Graesser, Lu et al., 2005; Graesser & Olde, 2003; Piaget, 1952). Cognitive equilibrium is restored after thought, reflection, problem solving and other effortful cognitive activities. When the learner is confused, there might be a variety of paths for the tutor to pursue. The tutor might want to allow the learner to continue being confused during the cognitive disequilibrium (and the affiliated increased physiological arousal that accompanies all affective states). The learner's self-regulated thoughts might hopefully restore equilibrium when the tutor

delays feedback to learner errors (Fox, 1993). Alternatively, after some period of time waiting for the learner to progress, the tutor might give indirect hints to nudge the learner into more productive trajectories of thought.

It is unclear what exactly should be the gold standard for deciding what emotions a learner is truly having. Should it be the learner, the expert, or an instrument? If it is the learner, what is the best manifestation of the emotions the learner is having? Would it be ratings, behavioral observations of the learner, physiological measures, or verbal reports? If it is an expert, what constitutes true expertise on accurate identification of an emotion? If it is an instrument, what is the most accurate instrument? Would it consist of physiological measures or non-intrusive sensing devices that classify emotions on the basis of facial expressions, body posture, speech, or natural language dialogue? We are uncertain about the best measures of affect states. However, we have explored a number of alternatives, such as trained judges coding learner emotions during learning (Craig, Graesser et al., 2004), trained judges coding videotapes of students learning (Graesser, McDaniel, Chipman, Witherspoon, & D'Mello, 2006), and non-intrusive sensing devices (D'Mello et al., 2005).

The present study pursues a new approach to detecting emotions of learners, called an *emote-aloud* procedure. Learners say out loud whatever emotions come to mind while interacting with the learning environment. Their expressions of emotions (called *emotes*) are classified into theoretical categories of emotions by trained experts. The emote-aloud procedure is analogous to the conventional *think aloud* protocols that are routinely incorporated in the methodologies of the cognitive and learning sciences (Ericsson & Simon, 1993; Graesser & Olde, 2003; Trabasso & Magliano, 1996). Participants are encouraged to focus on emotions in the emote-aloud procedure whereas any type of reflections (cognitive, social, emotional,

enactive) are encouraged in the think-aloud protocols. It should be noted that both emote-aloud and think-aloud methodologies are labor intensive because the protocols need to be transcribed, segmented into units, and classified into theoretical categories in addition to the normal statistical analyses. Therefore, researchers who use this type of methodology typically collect data from a small number of participants. Indeed, Newell and Simon (1972) collected data on less than a handful of participants in some of their classical studies of problem solving that used the think-aloud methodology.

The present project tracked the emotions that college students experience while interacting with AutoTutor, an intelligent tutoring system that helps students learn by holding a conversation in natural language (Craig, Driscoll, & Gholson, 2004; Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; Graesser, Person, & Harter, 2001; VanLehn, Graesser, Jackson, Jordan, Olney, & Rose, in press). AutoTutor was designed to simulate human tutors while it converses with students in natural language. AutoTutor begins by presenting a challenging question to the learner that requires about a paragraph of information to answer correctly. The typical response from the learner, however, is usually only one word to two sentences in length. Therefore, AutoTutor uses a series of pumps (“What else?”, “uh huh”), hints, short feedback and other dialogue moves to elicit responses from the learner that lead to a complete answer to the question. There are approximately 30 to 200 student and tutor turns, about the length of a dialogue with a human tutor, before the learner is able to give AutoTutor a paragraph of correct information. The conversational interactivity of AutoTutor makes it a good learning environment for exploring relations between emotions and characteristics of dialogue.

The general hypothesis is that there are systematic relations between these characteristics of dialogue and particular emotions that are manifested in the emote-aloud protocols (the emotes). More specifically, we investigated whether particular emotes (e.g., confusion, frustration, eureka/delight) are prevalent after AutoTutor's feedback (positive, neutral, negative), the directness of AutoTutor's dialogue moves (hints are less direct than assertions), the quality and verbosity of the learner's contributions, and the phase of the tutoring session. If these characteristics of dialogue are diagnostic in predicting emotions, then there is considerable hope in being able to use computers to automatically detect emotions in real time. That will be necessary in order to build an affect-sensitive AutoTutor.

#### *Emotions while Interacting with AutoTutor*

Previous observational studies have confirmed that a variety of emotions do in fact occur while college students interact with AutoTutor. Craig, Graesser et al. (2004) reported a study in which 5 trained judges observed six different affect states (confusion, frustration, boredom, flow/engagement, eureka, and neutral) that potentially occur during the process of learning introductory computer literacy with AutoTutor. The participants were 34 college students who had low subject matter knowledge about computer literacy according to a pretest (24 multiple choice questions). Trained judges recorded emotions that learners apparently were experiencing at random points during the interaction with AutoTutor, approximately every 5 minutes. Participants completed a pretest, interacted with AutoTutor for 30-45 minutes, and completed a posttest with multiple choice questions. Learning gains  $[(\text{posttest scores} - \text{pretest scores}) / (1.0 - \text{pretest scores})]$  were correlated with the incidence of these six emotions.

Craig, Graesser et al. (2004b) reported that there were significant correlations between learning gains and some of the emotions. Learning gains showed a significant positive

correlation with confusion ( $r = .33$ ) and flow/engagement ( $r = .29$ ), but a negative correlation with boredom ( $r = -.39$ ). Correlations with eureka ( $r = .03$ ) and frustration ( $r = -.06$ ) were near zero. The positive correlation between confusion and learning is consistent with a model that assumes that cognitive disequilibrium is one precursor to deep learning (Graesser & Olde, 2003; Graesser, Lu et al., 2005) and with models that help students learn how to overcome failure from getting stuck (Burlison & Picard, 2004). The findings that learning correlates negatively with boredom and positively with flow are consistent with predictions from Csikszentmihalyi's (1990) analysis of flow experiences. Experiences of eureka were much too rare; there was only one recorded eureka experience in 17 total hours tutoring among the 34 students. Frustration was also rarely experienced (only 3% of the recorded emotions), at least according to the expert judges. The percentage scores were higher for the affect states of confusion (7%), boredom (18%), and flow (45%).

The results of this initial observational experiment support a number of conclusions. First, there is a correlation between emotions and complex learning, although the causal relationship is not established. Second, there is a sufficient amount of affect during tutoring that it is feasible to study learning-affect relations in the context of AutoTutor. Third, confusion, flow/engagement, and boredom are frequent emotions that are apparent to judges in observational studies.

Graesser, McDaniel et al. (2006) conducted a follow-up study that had 28 college students learn computer literacy concepts for 32 minutes. These interactions were recorded on videotape for later analyses. The videotapes were stopped every 20 seconds for observers to judge the occurrence of emotions. The judges were either the self (the learner, immediately after the session with AutoTutor), a peer (another learner), or two expert judges (college students

trained on detecting emotions). We examined the proportion of judgments that were made for each of the emotion categories, averaging over the 4 judges. The most common affective states were confusion (.212), flow (.188), and boredom (.167); the remaining states of delight, frustration and surprise totaled .065 of the observations. When we inspected points when 2 or more judges were confident that there was an emotion, the most prominent affect state was confusion (.377), followed by delight/eureka (.192) and frustration (.191). Most of the time learners were either in a neutral state or in a subtle affective state (boredom or flow). The results of these two studies support the conclusion that it is important to monitor the following affective states in investigations of college students learning with AutoTutor: confusion, frustration, flow/engagement, boredom, and delight/eureka.

#### *AutoTutor's Mixed Initiative Dialog*

As mentioned earlier, AutoTutor is a fully automated computer tutor that simulates human tutors and holds conversations with students in natural language. The design of AutoTutor was inspired by explanation-based constructivist theories of learning (Alevan & Koedinger, 2002) and by previous empirical research that has documented the collaborative constructive activities that routinely occur during human tutoring (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Fox, 1993; Graesser & Person, 1994). AutoTutor helps students learn by presenting challenging questions from a curriculum script and engaging in a mixed-initiative dialog while the learner constructs an answer.

AutoTutor generates different categories of dialogue moves while interacting with the learner during the multi-turn interaction. AutoTutor provides *feedback* on what the student types in (positive, neutral, negative feedback), *pumps* the student for more information (“What else?”), *prompts* the student to fill in missing words, gives *hints*, fills in missing information with *assertions*, identifies and corrects *misconceptions* and erroneous ideas, *answers* the student’s

questions, and *summarizes* topics. A full answer to a question is eventually constructed during this dialog, which normally takes between 30 and 200 turns between the student and tutor (just as with human tutors).

AutoTutor's knowledge about the topic it is tutoring (computer literacy in this study) is represented by a curriculum script on the material and also Latent Semantic Analysis (LSA) (Foltz, 1996; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). LSA is a statistical technique that measures the conceptual similarity of any two texts that can range from 1 word to a lengthy article. LSA computes a geometric cosine (ranging from 0 to 1) that represents the conceptual similarity between the two text sources. In AutoTutor, LSA is used to assess the quality of student responses and to monitor other informative parameters, such as topic coverage and student ability level. The quality of the learner's responses is measured by comparing each response against two classes of content stored in the curriculum script: one that contains potential good answers to the topic being discussed (called *expectations*) and one that contains the anticipated bad answers (called *misconceptions*). The higher of the two geometric cosines (i.e., a measure of the conceptual match between student input and expectations/misconceptions) is considered the best conceptual match and determines how AutoTutor responds to the student contributions in the subsequent dialog turn (such as positive, neutral, or negative feedback). We have found our application of LSA to be quite accurate in evaluating the quality of learner responses (Graesser, Penumatsa, Ventura, Cai, & Hu, in press; P. Wiemer-Hastings, K. Wiemer-Hastings, & Graesser, 1999).

A session with AutoTutor is comprised of a set of subtopics (difficult questions or problems) that cover specific areas of the main topic (hardware, Internet, and operating systems). Each subtopic is covered by a series of turns in which AutoTutor maintains a conversation with

the student in an attempt to construct an answer to the current subtopic. When an acceptable answer, with the appropriate details, is gleaned from the learner's responses, AutoTutor moves to the next subtopic. At the end of each student turn, AutoTutor maintains a log file that captures the learner's response, a variety of assessments of the response, and the tutor's next move.

Table 1 provides an overview of the various channels of information in the student's interaction history. These are stored and extracted from AutoTutor's log files. A number of other information channels that are not relevant to this study do not appear in Table 1. The information channels can be divided into five categories: session information, learner verbosity, quality of learner contributions (LSA assessments), directness of AutoTutor in supplying information, and AutoTutor feedback. These measures are specified below.

\*\*\* INSERT TABLE 1 ABOUT HERE\*\*\*

*Session information.* These measure how far the learners have progressed through the AutoTutor session (a global index) or through a particular subtopic (a local index). The Subtopic Number indicates the number of questions that have been answered and covered in the 90-minute session. The Turn Number is a local measure of the number of student turns that attempt to answer a single question (subtopic). Intuitively, one would expect tiredness or boredom with a high Subtopic Number, and probably frustration with a high Turn Number because the student is stuck in the current subtopic.

*Learner verbosity.* Learner verbosity is the number of words or alphanumeric characters in the student's response. Short responses might reflect frustration or confusion. Long responses may reflect a deeper grasp of concepts, perhaps due to the experience of the learner being in a state of flow (Csikszentmihalyi, 1990).

Quality of learner contributions with Latent Semantic Analysis (LSA). The quality of learner contributions is evaluated by comparing the learners' contributions in each turn to good and bad answers. As discussed earlier, AutoTutor relies on LSA as its primary source of student response assessments. The local assessments for a given turn measure the student's response for that turn on the basis of its similarity to good versus bad answers (i.e., expectations versus misconceptions). The Local Good Score is the highest match to the set of expectations representing good answers. The Local Bad Score is the highest match to the set of bad answers. A high Local Good Score is indicative of progress, while a high Local Bad Score can be interpreted as a student's misconception. The Delta Local Good Score and the Delta Local Bad Score measure changes in the Local Good Score and the Local Bad Score, respectively, between the current turn N for a subtopic and turn N-1. Therefore, a large Delta Local Good Score might be associated with one of those rare *eureka* experiences.

The four Global parameters perform the same assessments as the Local parameters with the exception that the text used for the LSA match is an aggregation of all of the student's responses in a given subtopic, 1 through N. With this scheme, a student's past responses to a subtopic are considered in AutoTutor's assessment of his or her current response.

LSA-based assessments are not applied to all contributions of the learner within each turn. AutoTutor first segments the learners' verbal input within each turn into sentential units and classifies the units into categories. Some categories of learner contributions do not provide information that is relevant to an answer, such as short responses (*yes, okay*), meta-communicative statements (*What did you say?*), meta-comprehension statements (*I don't understand, that makes sense*), and learner questions. Other categories of learner contributions

are assertions that help answer the AutoTutor's main questions. It is the learner assertions that are analyzed with respect to the quality of learner contributions.

*Directness of AutoTutor in supplying information.* After the learner enters information within each turn, AutoTutor needs to generate the content of the next turn in a fashion that adapts to what the learner expressed in the previous turns and to the dialogue history. The content of most of AutoTutor's turns consists of short feedback (positive, negative, neutral) on the learner's contributions in turn N-1, one or more dialogue moves that help answer the question, and a final dialogue move that attempts to get the learner to contribute to the dialogue (such as asking the student a question). AutoTutor needs to generate the feedback and dialogue moves in a fashion that is pedagogically appropriate.

The dialogue moves generated by AutoTutor vary on a scale of "directness." At the low end of the continuum, AutoTutor provides pumps or hints to get the learner to do the talking and express answer information; this is advocated by constructivist theories and principles of active learning. At the high end of the continuum, AutoTutor delivers information through assertions and summaries. AutoTutor starts out giving indirect pumps and hints to get an expectation covered, but resorts to direct assertions when the learner has trouble articulating the expectation. AutoTutor starts out each main question (subtopic) by pumping the learner for information (e.g., *what else, uh huh*). After this pumping phase, AutoTutor identifies expectations that are not covered by the student and attempts to get these covered one expectation at a time. Whenever expectation E needs to be covered, AutoTutor launches a [hint → prompt → assertion] cycle in three successive AutoTutor turns. After the hint is given, the student sometimes articulates the answer correctly so AutoTutor exits the cycle and goes onto another expectation. If the student's response is inadequate, however, then AutoTutor presents a prompt on the next turn to get the

student to fill in a missing important word. If the student covers the expectation with an answer, AutoTutor goes on to the next expectation. If not, then AutoTutor generates an assertion in the next turn and thereby covers the expectation. This hint-prompt-assertion mechanism adapts to the learner's knowledge. AutoTutor ends up presenting mainly pumps and hints to students who are performing well, whereas low performing students require more prompts and assertions. The final phase of each main question is a summary answer, which is provided by AutoTutor.

The dialog moves chosen by AutoTutor can be regarded as an indicator of the amount of information delivered to the student. The five dialog moves presented in Table 1 can be mapped onto a scale in the following order: pump, hint, prompt, assertion, and summary. A pump conveys the minimum amount of information (on the part of AutoTutor) whereas a summary conveys the most amount of explicit information.

Within the context of the emote-aloud study, it is plausible that tutor directness may be correlated with the affect states of learners. For example, one might expect confusion to heighten after the occurrence of hints (when the student is expected to think, often to no avail) and to diminish in the presence of assertions and summaries (when the student can simply receive information from AutoTutor rather passively).

*Feedback.* AutoTutor's feedback is manifested in its verbal content, intonation, and other non-verbal conversational cues. Table 1 shows examples of AutoTutor's responses, characterized by the type of feedback provided. One could predict the occurrence of particular emotions as a result of the type of feedback provided. For example, repeated negative feedback could cause frustration in a motivated student, but boredom in a student lacking motivation.

The dialogue characteristics in AutoTutor's log files were mined on session information, learner verbosity, learner contribution quality, AutoTutor directness, and AutoTutor feedback

(see Table 1). These characteristics were assessed at points in the AutoTutor-learner dialogue when the learners expressed “emotes” in their emote-aloud protocols. We performed correlation analyses to investigate the relationship between the dialogue characteristics and the learner’s emotions.

## Methods

### *Participants*

The participants were 7 undergraduates in the department of psychology subject pool at the University of Memphis. Two participants were discarded because they rarely expressed any emotions; one expressed 6 emotes and the other 9 emotes during the entire 90-minute tutoring session. The remaining 5 learners expressed between 17 and 89 emotes in the tutoring sessions.

### *Materials*

*AutoTutor.* Participants interacted with AutoTutor for approximately 90 minutes on subtopics related to computer literacy (see Graesser et al., 2001 for detailed information about AutoTutor). Participants could cover up to 12 subtopics (questions) that required about a paragraph of information (3-7 sentences) in an ideal answer. The questions required answers that involved inferences and deep reasoning, such as *why*, *how*, *what-if*, *what if not*, *how is X similar to Y?* Although each question required 3-7 sentences (expectations) in an ideal answer, learners rarely give the complete answer in a single turn. A conversation typically occurs with multiple turns that take a few minutes.

The AutoTutor interface has 4 windows, as shown in Figure 1. Window 1 (top of screen) is the main question that stays on the computer screen throughout the conversation that involves answering the question. Window 2 (bottom of screen) is affiliated with the learner’s answer in any one turn and echoes whatever the learner types in via keyboard. Window 3 (left middle) is an

animated conversational agent that speaks the content of AutoTutor's turns. The talking head has facial expressions and some rudimentary gestures. Window 4 (right middle) is either blank or has auxiliary diagrams.

\*\*\* INSERT FIGURE 1 ABOUT HERE\*\*\*

Most turns of AutoTutor have three information slots (i.e., units, constituents). The first slot is feedback on the quality of the learner's last turn, as listed in Table 1. The second slot advances the conversation with either the dialogue moves in Table 1, corrections of misconceptions, or answers to student questions. The third slot is a cue for the floor to shift from AutoTutor as the speaker to the learner. This is accomplished by AutoTutor asking a question or gesturing for the learner to type in information. Discourse markers (*and also, okay, well*) connected the utterances generated from these three slots of information within a turn. The conversations managed by AutoTutor are sufficiently smooth that learners can get through the session with minimal difficulties.

*Knowledge tests.* Two 24-item multiple-choice tests on the domain of computer literacy were used to assess prior domain knowledge and learning gains. There were two questions associated with each of the 12 subtopics. The tests were counterbalanced as pretest and posttest. However, learning gains were not relevant to this study so these tests were not analyzed for the current paper.

### *Procedure*

As participants came into the lab, they were given an informed consent followed by a pretest. Then the participants interacted with AutoTutor for approximately 90 minutes, during which they engaged in an emote-aloud activity. The participants were videotaped during the interaction with AutoTutor. They were asked to make verbal reports when they experienced an

affective state. Participants were given a list with eight affective states along with definitions. The list of affective states consisted of anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration.

The affective states were functionally defined for the participants, based on a dictionary. *Anger* was defined as a strong feeling of displeasure and usually of antagonism. *Boredom* was defined as the state of being weary and restless through lack of interest. *Confusion* was defined as a failure to differentiate similar ideas or to relate ideas. *Contempt* was defined as the act of despising, with a lack of respect or reverence for something. *Curious* was defined as an active desire to learn or to know. *Disgust* was defined as marked aversion aroused by something highly distasteful. *Eureka* was defined as a feeling used to express triumph on a discovery. *Frustration* was defined as a feeling of making vain or ineffectual all efforts however vigorous; a deep chronic sense or state of insecurity and dissatisfaction arising from unresolved problems or unfulfilled needs. After the 90 minute session ended, a posttest was administered, followed by a debriefing.

#### *Selection and coding of emote expressions*

*Coding of emotes.* The videotapes of emote-aloud protocols and interactions with AutoTutor were transcribed and analyzed for occurrences of emotion expressions (emotes). Two trained judges recorded all emotes that were expressed by participants in the categories of anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration. Inter-judge reliability scores (Cohen's kappa) were nearly perfect in making these judgments because the judges relied on explicit matches to the labels presented to the learners during the instructions (i.e., anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration).

*Data selection.* There were a small number of occurrences of anger ( $n = 17$ ), contempt ( $n=8$ ), curiosity ( $n=1$ ), and disgust ( $n=5$ ), so these were not included in subsequent analyses. This data cleaning procedure resulted in reliable data only for boredom (40), confusion (53), eureka (28), and frustration (49).

The AutoTutor log files were mined to obtain data on the various dialog channels presented in Table 1. More specifically, the turn that immediately preceded the emote-aloud was selected as the representative turn for that emote. If any of the 22 dialog parameters for such a turn were missing, that turn and the associated emotes were discarded from the analysis. This resulted in a further reduction in the database to 145 records.

### Results and Discussion

Information from AutoTutor's log files were correlated with the emotes expressed by the learners. Significant Pearson correlations were found for confusion, eureka, and frustration, whereas no significant correlations were found for boredom. All reported correlations in this section were based on 145 observations ( $df = 143$ ) and reached a significance level of at least  $p < .05$  on a two tailed test. These analyses will be segregated for confusion, eureka, and frustration.

#### *Correlations with Confusion*

Table 2 presents significant correlations between confusion and AutoTutor's dialog channels. The negative correlations in Table 2 can be readily explained. A high Delta Global Good Score is an indicator that the learner is understanding the material, so it is negatively correlated with confusion. A student's confusion is lower after AutoTutor presents an assertion; the tutor states facts and minimally engages the student in thought. As would be expected, confusion is negatively correlated with positive feedback. When the learner is confused, the

learner's answers are generally not very accurate and positive feedback should decrease. During confusion, the student is in a state of cognitive disequilibrium.

\*\*\* INSERT TABLE 2 ABOUT HERE\*\*\*

The positive correlations in Table 2 can also be meaningfully interpreted. Confusion is positively correlated with the presence of a hint, neutral negative feedback, and neutral feedback. The confusion that is experienced when the participant is presented with a hint could be attributed to the higher level of thought required to answer the hint, in the state of cognitive disequilibrium. The hint presents the learner a small bit of information and encourages him/her to follow a thread of reasoning to construct the answer. If the learner does not follow the correct path, then the learner runs the risk of becoming confused from the lack of synchrony between the hint and the learner's own mindset. This confusion is also reflected in the accompanying neutral negative and neutral feedbacks. The neutral or neutral negative feedback may present confusion when the feedback is not compatible with the learner's expectations on their own level of understanding.

#### *Correlations with Eureka*

The correlations with eureka in Table 2 are relatively easy to interpret. In general, more eureka occurs as the user's LSA scores for good answers increase. This would be an indication that users are learning the material. Eureka also shows a strong relationship to both negative feedback and positive feedback, indicating that eureka can be influenced by AutoTutor's feedback. These findings support the generalization that eureka occurs when there is a synchrony between the learner's knowledge and feedback from AutoTutor.

In retrospect, our labeling of this emotion as eureka is most likely to be misleading. This emotion was functionally a form of delight after the learner gave a correct answer – not a full

eureka experience. The initial instructions mentioned eureka, a colorful rare term, which learners liberally incorporated in their emote-aloud protocols when they were experiencing the delight of giving a correct answer. A bona fide eureka experience would consist of a flash of insight about a difficult problem, followed by extreme joy. True eureka experiences are much more infrequent than our data suggest (Craig, Graesser et al., 2004).

### *Correlations with Frustration*

Table 2 presents significant correlations between frustration and AutoTutor's dialog characteristics. The correlation between assertions and frustration was small but significant. As expected, frustration is positively correlated with negative feedback and negatively correlated with positive feedback. Frustration occurs when there is a lack of synchrony between AutoTutor's feedback and the learner's views on his/her own understanding.

One possible reason that assertions tend to trigger frustration is because the learner's understanding is so poor that AutoTutor must resort to delivering the information to the learner (a last resort). Assertions are given after AutoTutor tries to get the learner to articulate the information through pumps, hints, and prompts. Thus, the lack of understanding on the part of the learner leads to frustration as AutoTutor ends up needing to merely deliver the information.

### General Discussion

It appears that there are significant relationships between the content of dialog and the emotions experienced during learning. The current study found significant correlations between dialog and the affective states of confusion, eureka (or perhaps better viewed as delight), and frustration. These emotional states were driven by characteristics of AutoTutor's feedback, the directness of AutoTutor's dialogue moves, and the quality of the learner's contributions. In contrast, emotions were not affected by the verbosity of the learners' contributions and very little

by the phase of the tutoring session (beginning, middle, versus end) and the phase of covering the answer to the main questions.

One theoretical explanation of the links between emotions and learning would appeal to a *grounding criterion*. A collaborative theory of communication (Schober & Clark, 1989) assumes that speech participants need to converge on a mutual belief that they understand each other to a criterion sufficient for the current purposes. The grounding criterion is met during delight/eureka, when the student's contributions are high in quality and therefore AutoTutor gives positive feedback rather than negative feedback. There is a breakdown in communication when this grounding criterion is not met. That is, there is a state of cognitive disequilibrium (Graesser, Lu et al., 2005; Otero & Graesser, 2001), which produces confusion (Craig, Graesser et al., 2004b) and sometimes frustration. States of confusion and frustration result in negative feedback rather than positive feedback from the Tutor. According to the data, confusion is often prompted by AutoTutor's indirect hints rather than its direct assertions, whereas frustration occurs when the grounding criterion is so far from being met that AutoTutor ends up delivering the correct information with assertions. It is apparent that the links between cognition and emotion can be quite complex, interactive, and subtle.

The affective state of boredom was not significantly predicted by patterns of dialogue. However, it is tempting to speculate how boredom may result from problematic dialogue. If the learner fails to reach a grounding criterion, then the learner eventually gives up and disengages, resulting in boredom. In contrast, frustration occurs when the tutor moves ahead before the learner has reached understanding. The grounding criterion may be restored in a state of understanding, which is occasionally preceded by an abrupt transition of eureka (quick insight).

These correlations were supportive of a grounding criterion hypothesis of dialog and emotions. However, more conclusive testing must be conducted to further test the validity of these links. If the grounding criterion holds, then it would give indications of how to help structure a dialog to generate the best interaction between emotions and learning. For the most part, it appears that the emotions experienced during interactions with AutoTutor are based on dialog moves moving on two dimensions: Feedback and directness. Our future research will investigate further the effects of these two scales on the emotions experienced during AutoTutor interactions.

One technological challenge lies in identifying what sensing devices and automated emotion classifiers we should integrate with AutoTutor. An automated affect classifier is of course needed to make AutoTutor responsive to learner emotions. We have collected some data that record the dialogue history, facial action units, positions of the learner's body, and other sensory channels during learning with AutoTutor. There are systematic relations between these sensing channels and particular emotions (D'Mello, Craig, Gholson, Franklin, Picard, & Graesser, 2005; Kapoor & Picard, 2005). The present study has documented relations between dialogue and emotions, but there are other sensing channels that are potential diagnostic of learning emotions. For example, particular facial expressions are correlated with particular emotions (Craig, D'Mello, Witherspoon, Sullins, & Graesser, 2004). Frustration is associated with outer brow raise, inner brow raise, and the dimpler whereas confusion is associated with brow lowerer, lid tightener, and lip corner puller. Posture may be correlated with interest (Mota & Picard, 2003). Students experiencing flow may tend to lean forward in the chair whereas the bored students either slump back or are persistently agitated. If we record speech, then affective states may be induced from a combination of lexical, acoustical, and prosodic features (Litman &

Forbus-Reilly, 2004). We believe that most of these features from the various modalities can be detected in real time automatically on computers. Whether an automated affect detector can be achieved awaits future research and technological development.

### Acknowledgments

We thank our research colleagues in the Emotive Computing Group and the Tutoring Research Group (TRG) at the University of Memphis (<http://www.autotutor.org>). We gratefully acknowledge our partners at the Affective Computing Research Group at MIT. This research was supported by the National Science Foundation (REC 0106965 and ITR 0325428) and the DoD Multidisciplinary University Research Initiative administered by ONR under grant N00014-00-1-0600. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF, DoD, or ONR.

## References

- Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer based Cognitive Tutor. *Cognitive Science*, 26, 147-179.
- Bower, G.H. (1981). Mood and memory. *American Psychologist*, 36, 129-148.
- Burleson, W., & Picard, R.W. (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. (2004). In *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems: Workshop on social and emotional intelligence in learning environments*. Maceio-Alagoas, Brazil: ITS.
- Carver, C. S. (2004). Negative affects deriving from the behavioural approach system. *Emotion*, 4, 3-22.
- Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471-533.
- Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence*, 16, 555-575.
- Craig, S.D., D'Mello, S., Witherspoon, A., Sullins, J., & Graesser, A.C. (2004). Emotions during learning: The first step toward an affect sensitive intelligent tutoring system. In the *Proceedings of the International Conference on eLearning*. Boston, MA: AACE.
- Craig, S. D., Driscoll, D., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia*, 13, 163-183.
- Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, 29, 241-250.

- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper-Row: NY.
- Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Orlando, FL:Harcourt.
- De Vicente, A., & Pain, H. (2002). Informing the detection of students' motivational state : An empirical study. In S.A. Cerri, G. Gouarderes, and F. Paraguacu (Eds *Proceedings of the sixth international conference on intelligent tutoring systems* (pp.933-943). Berlin, Germany: Springer.
- Deci, E. L., & Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
- D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International conference on Intelligent User Interfaces* (pp. 7-13) New York: AMC Press
- Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
- Ekman, P. (2003). *Emotions Revealed*. New York: Times Books.
- Ekman, P, & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Revised edition. Cambridge, MA: The MIT Press.

- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, and Computers*, 28, 197-202.
- Fox, B. (1993). *The human tutorial dialogue project*. Hillsdale: Erlbaum.
- Fredrickson, B. L. and C. Branigan. 2005. Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion*, 19, 313-332.
- Gee, J.P. (2003). *What video games have to teach us about language and literacy*. New York: Macmillan.
- Graesser, A.C., Chipman, P., Haynes, B.C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*. 48, 612-618.
- Graesser, A.C., Jackson, G.T., & McDaniel, B. (in press). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology*.
- Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180-193.
- Graesser, A.C., Lu, S., Olde, B.A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, 33, 1235-1247.
- Graesser, A.C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S. (2006). Detection of emotions during learning with AutoTutor. In the *Proceedings of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

- Graesser, A. C. & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95, 524–536.
- Graesser, A.C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (in press). Using LSA in AutoTutor: Learning through mixed initiative dialogue in natural language. In T. Landauer, D. McNamara, S. Dennis, and W. Kintsch (Eds.), *LSA: A Road to meaning*. Mahwah, NJ: Erlbaum.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.
- Graesser, A. C., Person N., Harter, D. & The Tutoring Research Group (2001) Teaching tactics and dialog in AutoTutor, *International Journal of Artificial Intelligence in Education*, 12, 257–279.
- Graesser, A.C., Person, N., Lu, Z., Jeon, M.G., & McDaniel, B. (2005). Learning while holding a conversation with a computer. In L. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 143-167). Greenwich, CT: Information Age Publishing.
- Guhe, M., Gray, W. D., Schoelles, M. J., & Ji, Q. (2004, July). *Towards an affective cognitive architecture*. Poster session presented at the Cognitive Science Conference, Chicago, IL.
- Kapoor, A., & Picard, R. (2005). Multimodal affect recognition in learning environments” *ACM Multimedia*.
- Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion. In T. Okamoto, R. Hartley, Kinshuk, & J. P. Klus (Eds.), *Proceedings IEEE International*

- Conference on Advanced Learning Technology: Issues, Achievements and Challenges* (pp.43-48). Madison, Wisconsin: IEEE Computer Society.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.
- Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257-307). San Diego, CA: Academic Press.
- Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.
- Linnenbrink, E. A., & Pintrich, P. R. (2002). The role of motivational beliefs in conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 115-135). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42<sup>nd</sup> annual meeting of the association for computational linguistics* (pp. 352-359). East Stroudsburg, PA: Association for Computational Linguistics.
- Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. In *proceedings of the human language technology conference: 3<sup>rd</sup> meeting of the North*

- American chapter of the association of computational linguistics* (pp. 52-54). Edmonton, Canada: ACL.
- Mandler, G. (1976). *Mind and emotion*. New York: Wiley.
- Metcalfe, J., & Kornell, N. (2005). A region or proximal of learning model of study time allocation. *Journal of Memory and Language*, 52, 463-477.
- Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist*, 37, 107-114.
- Mota, S. & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. *Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI*, June, 2003.
- Newell, A., and H. A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. New York: Cambridge University Press.
- Otero, J., & Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction*, 19, 143-175.
- Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.
- Picard, R. W. (1997). *Affective computing*. Cambridge, Mass: MIT Press.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.
- Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.

Schutzwohl A, and K. Borgstedt. (2005). The processing of affectively valenced stimuli: The role of surprise. *Cognition & Emotion, 19*, 583-600

Stein, N. L., & Levine, L. J. (1991). Making sense out of emotion. In W. Kessen, A. Ortony, & F. Kraik (Eds.), *Memories, thoughts, and emotions: Essays in honor of George Mandler* (pp. 295-322). Hillsdale, NJ: Erlbaum.

Trabasso, T., & Magliano, J. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 225-286.

VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., & Rose, C.P. (in press). When are tutorial dialogues more effective than reading? *Cognitive Science*.

Vorderer, P. (2003). Entertainment theory. In B. Jennings, D. Roskos-Ewoldsen, & J. Cantor (Eds.), *Communication and emotion: Essays in honor of Dolf Zillmann* (pp. 131-153). Mahwah, NJ: Erlbaum.

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. C. (1999). Improving an intelligent tutor's comprehension of students with latent semantic analysis. *International Journal of Artificial Intelligence in Education, 535-542*. Amsterdam: IOS Press.

Table 1: *Description of the Information Mined from AutoTutor's Log Files at the End of Each Student Turn.*

Channel	Sub channel	Description
Session information	Subtopic Number	The current subtopic (question) in this session
	Turn Number	The number of the conversation turn within a subtopic
Learner verbosity	Number of words	The number of words in the student's turn
	Number of characters	The number of characters in the student's turn
Quality of learner contributions (Latent Semantic Analysis assessments)	Local Good Score	Similarity of content of student's turn to an expectation
	Delta Local Good Score	The change in the Local Good Score
	Global Good Score	Similarity of the history of student turns to expectations
	Delta Global Good Score	The change in the Global Good Score
	Local Bad Score	Similarity of content of student's turn to a bad answer
	Delta Local Bad Score	The change in the Local Bad Score
Directness of AutoTutor in supplying information	Global Bad Score	Similarity of the history of student turns to bad answers
	Delta Global Bad Score	The change in the Global Bad Score
	Pump	Minimal information provided. e.g. "What else"
	Hint	Provides a hint to the student to fill in proposition
	Prompt	Prompts student to fill in a missing content word
AutoTutor feedback	Assertion	Asserts information about an expectation
	Summary	Provides a summary of the answer
	Positive	Provides feedback terms such as: "good job", "correct"
	Neutral Positive	Provides feedback terms such as: "yeah", "hmm right"
	Neutral	Provides feedback terms such as: "uh huh", "alright"
Neutral Negative	Provides feedback terms such as: "possibly", "kind of"	
Negative	Provides feedback terms such as: "wrong", "no"	

Table 2: *Significant correlations between dialogue characteristics and confusion, eureka, and frustration.*

Channel	Sub channel	Confusion	Eureka	Frustration
Session information	Subtopic Number		.23	
	Turn Number			
Learner verbosity	Number of words			
	Number of characters			
Quality of learner contributions (Latent Semantic Analysis assessments)	Local Good Score	-.19	.26	
	Delta Local Good Score		.19	
	Global Good Score		.24	
	Delta Global Good Score		.20	
	Local Bad Score			
	Delta Local Bad Score			
	Global Bad Score			
	Delta Global Bad Score			
Directness of AutoTutor in supplying information	Pump	.17		
	Hint			
	Prompt	-.22		
	Assertion			
Summary				
AutoTutor feedback	Positive	-.27	.51	-.33
	Neutral Positive			
	Neutral	.20		
	Neutral Negative	.18		
	Negative		-.41	.41

Figure 1: Interface of AutoTutor

How is the packet switching model of message transmission like the postal system?

Packets are sent to intermediate destinations before being routed to their final destination.

Start | C:\AutoTutor | ts-os-nold.param -... | C:\WINDOWS\Sy... | 12:05 PM