Emotions during the Learning of Difficult Material

Arthur C. Graesser
University of Memphis

Sidney D’Mello
University of Notre Dame

Psychology of Learning and Motivation, volume 57,
Edited by Brian Ross

Art Graesser
Department of Psychology & Institute for Intelligent Systems
202 Psychology Building
University of Memphis
Memphis, TN 38152-3230
901-678-4857
901-678-2579 (fax)
a-graesser@memphis.edu

Sidney D’Mello
Department of Computer Science
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556
Phone:(901)-[378-0531]
http://www.nd.edu/~sdmello
Email: sdmello@nd.edu

Keywords: Conversational agents, Cognitive Disequilibrium, Emotions, Intelligent Tutoring Systems, Learning, Tutoring

17,646 words
# Table of Contents

1. Introduction  
1.1 Perspective on Emotions  
1.2 Complex Learning  
1.3 A Cognitive Disequilibrium Theoretical Perspective  

2. Learning Materials and Tasks  
2.1 Learning Environments with Pedagogical Agents  
   2.1.1 AutoTutor  
   2.1.2 Operation ARIES!  

3. Emotions that Occur during Difficult Learning Materials and Tasks  
3.1 What Emotions Occur during Complex Learning?  
   3.1.1 Trained Observers  
   3.1.2 Emote Aloud Protocols  
   3.1.3 Identification of Emotions by Learners, Peers, Trained Judges and Expert Teachers  
   3.1.4 Comparisons of Different Computer-Based Learning Environments  
3.2 Temporal Dynamics of Emotions  
   3.2.1 Duration of Emotions  
   3.2.2 Transitions between Emotions  

4. Responding to and Eliciting Student Emotions  
4.1 Emotion-Sensitive AutoTutor  
4.2 Planting Cognitive Disequilibrium  

5. Conclusions  

6. Acknowledgements  

7. References  

Tables  

Figure Caption Page  

Figures
Abstract

Students experience a variety of emotions (or cognitive-affective states) when they are assigned difficult material to learn or problems to solve. We have documented the emotions that occur while college students learn and reason about topics in science and technology. The predominant learning-centered emotions are confusion, frustration, boredom, engagement/flow, curiosity, anxiety, delight, and surprise. A cognitive disequilibrium framework provides a reasonable explanation of why and how these emotions arise during difficult tasks. The student is in the state of cognitive disequilibrium when confronting impasses and obstacles, which launches a trajectory of cognitive-affective processes until equilibrium is restored, disequilibrium is dampened, or the student disengages from the task. Most of our work has been conducted in computerized learning environments (such as AutoTutor and Operation ARIES!) that help students learn with pedagogical agents that hold conversations in natural language. An emotion-sensitive AutoTutor detects student emotions and adaptively responds in ways to enhance learning and motivation.
Emotions during the Learning of Difficult Material

1. Introduction

Emotions undoubtedly play a central role in linking learning and motivation. This is quite obvious when students struggle with complex technical texts, challenging writing assignments, and difficult problems to solve in their courses. The highly motivated students have the persistence to complete the expected tasks and experience positive emotions when the tasks are successfully accomplished. They experience curiosity when the topics interest them, eureka moments when there are deep insights and discoveries, delight when challenges are conquered, and intense engagement to the point where time and fatigue disappear. However, in route to their achieving these goals and their experiencing these positive affective states, they experience a rough terrain of confusion, frustration, and other negative emotions as they confront various obstacles in comprehension, production, reasoning, and problem solving. So there is a mixture of positive and negative affective states during the moment-to-moment process of learning. The students with low motivation and little interest in the material experience much more negative emotions than positive emotions. They quickly become bored and disengage after encountering a small amount of obstacles and dense technical content.

Moment-to-moment emotions both reflect and influence learning so it is important to understand the emotion dynamics that accompany complex learning. However, until recently, researchers rarely investigated emotion dynamics during learning at a fine-grain level. The primary goal of this chapter is to clarify the role of emotions during the process of learning difficult material. The chapter documents the learner-centered emotions that occur in a number of learning environments that cover difficult content and require complex reasoning. The occurrence, duration, and sequencing of these emotions has been investigated in an intelligent
tutoring system (AutoTutor) that helps students learn by holding a conversation in natural language. We propose a theoretical framework to account for these distributions of emotions during complex learning. According to a cognitive disequilibrium model, the learner is in the state of cognitive disequilibrium when confronting obstacles, which launches a trajectory of cognitive-affective processes until equilibrium is restored, disequilibrium is dampened, or the student disengages from the task. The chapter also describes an emotion-sensitive AutoTutor that detects student emotions and that generates discourse moves with affect-sensitive expressions designed to scaffold deeper learning and motivation.

1.1 Perspectives on Emotion

Contemporary psychological theories routinely assume that emotion and cognition are tightly integrated rather than being loosely linked systems (Bower, 1992; Clore & Huntsinger, 2007; Isen, 2008; Lazarus, 2000; Lewis, Haviland-Jones, & Barrett, 2008; Mandler, 1984, 1999; Ortony, Clore, & Collins, 1988; Picard, 1997; Scherer, Schorr, & Johnstone, 2001; Stein, Hernandez, & Trabasso, 2008). However, the learning of difficult material has rarely been the focus of cognitive research. Researchers have instead concentrated on paradigms that examine links between emotions and perception, memory, causal attribution, decision making, creative problem solving, and mental deliberation. Moreover, the emotions that researchers have investigated are not the typical emotions that students experience during the learning of difficult material, as will be conveyed throughout this chapter. Instead, most of the emotion research has targeted the six “basic” emotions investigated by Ekman (Ekman, 1992) that are readily manifested in facial expressions: sadness, happiness, anger, fear, disgust, and surprise. Ekman’s big six emotions do not frequently occur during the learning sessions of relevance to this chapter, except for an occasional occurrence of surprise. For this reason, we believe it is time for
researchers investigating emotion to emancipate themselves from restricting their focus to Ekman’s big 6 emotions and to becoming more open to a broader range of emotions and contexts -- in our case the learning of difficult material.

With rare exception, the psychological research investigating links between emotions and complex learning have not examined moment-to-moment dynamics of emotions. Instead, the goal of many researchers has been to identify traits of students that persist over time and tasks. Measures of enduring traits of relevance to learning have tapped constructs of motivation, self-concept, and goal orientations (Boekaerts, 2007; Daniels, et al., 2009; Frenzel, Pekrun, & Goetz, 2007; Linnenbrink, 2007; Pekrun, Elliot, & Maier, 2006; Schutz & Pekrun, 2007). For example, students vary in the extent to which they are an academic risk takers who are not afraid of negative feedback versus cautious learners who prefer safe tasks that lead to positive feedback (Clifford, 1988; Meyer & Turner, 2006). Students vary on the extent to which they are mastery-oriented versus performance-oriented and whether they avoid tasks that elicit negative emotions (Deci & Ryan, 2002; Pekrun, Elliot, & Maier, 2006). Some students have the self-concept that they are good or bad at particular topics (e.g., math, physics, literature) so they do not try to master the material whereas others believe effort devoted to any topic will lead to eventual mastery (Dweck, 2002). Intrinsically motivated learners derive pleasure from the task itself (e.g., enjoyment from problem solving), while learners with extrinsic motivation rely on external rewards (e.g., receiving a good grade). Learners with more intrinsic motivation display greater levels of pleasure, more active involvement in tasks (Harter, 1992; Tobias, 1994), more task persistence with lower levels of boredom (Miserandino, 1996), and less anxiety and anger (Patrick, Skinner, & Connell, 1993). These allegedly enduring traits are expected to
systematically mediate moment-to-moment emotional experience but they do not directly account for the dynamics of emotions during complex learning.

Research on tutoring has uncovered several discoveries about the relations between emotions and the learning of difficult material. Moment-to-moment emotions have been tracked in the context of human-to-human tutoring (Lehman, Mathews, et al., 2008) and computer-to-human tutoring (Arroyo et al., 2009; Baker, D'Mello, Rodrigo, & Graesser, 2010; Kapoor, Burleson & Picard, 2007; Calvo & D’Mello, 2010; Conati & Maclaren, 2010; D'Mello & Graesser, 2012, in press; Litman & Forbes-Riley, 2006; McQuiggan, Robison, & Lester, 2010). These tutoring sessions typically last 30 minutes to 2 hours and cover challenging content and skills. The computer-to-human tutoring has an advantage to the extent there can be systematic control over the presentation of materials, assessment of student progress, and strategies of interacting with the student. Therefore, this chapter will concentrate on the research on emotions that involve computer-to-human tutorial interaction.

It is appropriate to define our conception of emotion in the context of the research discussed in this chapter. We decided to not be overly constrained at this point in our explorations of learning-emotion connections because the field is at its infancy in understanding these connections. According to our definition, emotions are complex configurations of social-cognitive-affective-behavioral-physiological states that dynamically unfold over time in complex context-sensitive ways that sometimes defy the ascription of simple labels (e.g., frustration, boredom, surprise, etc.). What counts as an emotion in the present context is any social-affective state that noticeably deviates from a neutral base state. That being said, we nevertheless will use emotion labels for the heuristic purpose of communicating our findings at this early discovery phase of research. We acknowledge that the fact that we have a word, label, or phrase to
describe an emotion does not mean that we should reify it to the status of a scientific construct (Graesser & D’Mello, 2011). The words we use to describe emotions are products of folklore, the historical evolution of the language, the social context of interpretation, and other cultural fluctuations that are guided by principles very different from scientific theories of psychological mechanisms. This view is accepted by contemporary theories of emotions that differentiate the fundamental psychological dimensions of valence (a bad to good continuum) and intensity (low to high arousal) from the folklore, labels, and contextual interpretations of emotions (Barrett, 2006; Russell, 2003). The labels we use for emotions should therefore be considered pre-theoretical labels that serve our heuristic need to communicate some basic empirical findings at this early phase of research.

1.2 Complex Learning

This chapter focuses on connections between emotion and cognition that are prevalent during complex learning when students encounter difficult material. Complex learning occurs when a person tries to understand technical texts, to reason with multiple sources of information, to solve challenging problems, and to resolve conflicts. For example, complex learning occurs when a person attempts to comprehend a legal document, to locate a restaurant in a new city, to fix a broken piece of equipment, or to decide whether to purchase a new home. Comprehension, reasoning, and problem solving normally require effortful reflection and inquiry because there is a discrepancy between (a) the immediate situation and (b) the person's knowledge, skills, and strategies. The person is in the state of cognitive disequilibrium which launches a trajectory of social-cognitive-affective-behavioral-physiological states and processes until equilibrium is restored, disequilibrium dies out, or the person disengages from the task. A theoretical model is
articulated later in this section that specifies the trajectories of emotions that occur when people experience cognitive disequilibrium.

Complex learning is different from learning that is less taxing on the cognitive system. A distinction is sometimes made between tasks that involve shallow versus deep levels of cognitive processing, with a continuum of depth levels that was defined by Bloom over 50 years ago (1956). The major categories in Bloom’s original taxonomy are presented below:

1. **Recognition.** The process of verbatim identification of specific content (e.g., terms, facts, rules, methods, principles, procedures) that was explicitly presented in the learning material.

2. **Recall.** The process of actively retrieving from memory and producing content that was explicitly mentioned in the learning material.

3. **Comprehension.** Demonstrating conceptual and inferential understanding of the learning material by interpreting, paraphrasing, translating, explaining, or summarizing information.

4. **Application.** The process of applying knowledge extracted from the learning material to a problem, situation, or case (fictitious or real-world) that was not explicitly expressed in the learning material.

5. **Analysis.** The process of decomposing elements and linking relationships between elements.

6. **Synthesis.** The process of assembling new patterns and structures, such as constructing a novel solution to a problem or composing a novel message to an audience.
(7) **Evaluation.** The process of judging the value or effectiveness of a process, procedure, or entity, according to some criteria and standards.

The cognitive processes tend to be progressively more difficult with higher numbers, although differences among levels 4-7 are difficult to scale ordinally.

A more recent system contrasts three levels of learning events that vary in cognitive complexity (Koedinger, Corbett, & Perfetti, in press). They are:

1. **Memory and Fluency-building processes.** Processes involved in strengthening memory, compiling knowledge, and producing more automatic and composed (“chunked”) knowledge.

2. **Induction and Refinement processes.** Processes that improve the accuracy of knowledge, such as focused perception, generalization, discrimination, classification, categorization, and schema induction.

3. **Understanding and Sense-making processes.** Processes involved in explicit understanding and reasoning, such as comprehension of verbal descriptions, explanation-based learning, scientific discovery, and rule-mediated deduction.

The third level is most challenging and may require collaborative social interactions with experts in order to master difficult material with sufficient depth and accuracy. The complex learning of relevance in this chapter consists of this level 3 rather than levels 1 and 2.

**1.3 A Cognitive Disequilibrium Theoretical Perspective**

Cognitive disequilibrium plays a central role in the theoretical framework we have adopted in our research on moment-to-moment emotions during complex learning. The cognitive disequilibrium framework postulates an important role of impasses and discrepancies during complex learning. Cognitive disequilibrium is a state that occurs when people face
obstacles to goals, interruptions, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (D'Mello & Graesser, 2012, in press; Festinger, 1957; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Graesser & Olde, 2003; Otero & Graesser, 2001; Mandler, 1984, 1999; Piaget, 1952; Schwartz & Bransford, 1998; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003; Stein et al., 2008). There is a salient discrepancy between the student’s knowledge and the demands of the immediate situation. There are obstacles and interruptions to the goals that the student is trying to achieve. Cognitive disequilibrium triggers some important learner-centered emotions and also inquiry (e.g., exploration, question asking). Cognitive equilibrium is restored after thought, reflection, problem solving, information search, and other effortful deliberations.

This cognitive disequilibrium framework postulates that the complex interplay between external events that trigger impasses, discrepancies, and the resulting cognitive disequilibrium are the key to understanding the emotions that underlie complex learning. For example, confusion and sometimes frustration are likely to occur during cognitive disequilibrium. When the cognitive disequilibrium persists, there is the risk of the student eventually disengaging from the task and experiencing boredom. When the challenges of cognitive disequilibrium are conquered, then the student experiences the positive emotions of delight or flow/engagement (Csikszentmihalyi, 1990). Students are in a state of flow (Csikszentmihalyi, 1990) when they are so deeply engaged in learning the material that time and fatigue disappear. The zone of flow occurs when the structure of the learning environment matches a student’s zone of proximal development (Brown, Ellery, & Campione, 1998; Vygotsky, 1978); the student is presented with just the right sort of materials, challenges, and problems to the point of being totally absorbed. Flow occurs when there is an optimal oscillation between cognitive disequilibrium and the resolution of the disequilibrium. The parameters of this oscillation vary among students. Some
students can accommodate or even enjoy high levels of disequilibrium, confusion, and frustration over a lengthy time span. Some games engineer the parameters to optimize the engagement and flow.

Figure 1 conveys the essence of how the cognitive disequilibrium framework explains emotions during complex learning (D’Mello & Graesser, 2012, in press). Students start out in a state of equilibrium, engagement, and possibly flow. Then some event, stimulus, or thought occurs that creates an impasse and the student experiences confusion (and sometimes surprise when the phenomenal experience is abrupt). If the confusion is resolved, the student returns to engagement and flow, thereby completing a cycle of oscillation (and sometimes delight when a difficult impasse is conquered). Alternatively, if the impasse is not resolved and the goal is persistently blocked, then the student experiences frustration. As the student struggles with the frustration, there invariably will be additional impasses and resulting confusion. At some point the persistent failure will lead to disengagement and boredom. Boredom can also lead to frustration if the student is forced to endure the learning session after mentally disengaging. It will take considerable effort to conquer the boredom and frustration, a willful activity that many students do not pursue.

The cognitive disequilibrium framework can account for transitions between emotions and how the emotions interact with events and cognitive states. However, the duration of the emotions and the likelihood of taking the transitions depend on a host of factors, such as the student’s traits and knowledge, the timing of the cognitive processes, the importance of the task goals, the difficulty of the tasks, and other factors. Consider some of the examples below.
(1) The student is mastery-oriented with a high degree of conscientious and persistence. The student will spend more time in the states of confusion and frustration before disengaging and experiencing boredom.

(2) At some point during confusion, an impasse is suddenly resolved in a flash of insight (eureka) so the student experiences delight in route to a return to flow.

(3) A person’s automobile breaks down in the middle of the night so the person spends hours of frustration reading the drivers’ manual and experiencing new cycles of confusion with new forms of impasse.

(4) A student is an academic risk taker so the student can experience flow in the face of a large number of impasses, setbacks, and negative feedback.

(5) A student’s self-concept of math aptitude is low so the student quickly becomes bored and disengages when given challenging math problems.

Mood states may also mediate emotions and complex learning. Mood theories highlight the important role of baseline mood states (positive, negative, or neutral) on learning, particularly for creative problem solving. Flexibility, creative thinking, and efficient decision-making in problem solving have been linked to experiences of positive affect (Clore & Huntsinger, 2007; Fielder, 2001; Fredrickson & Branigan, 2005; Isen, Daubman, & Nowicki, 1987), whereas negative affect has been associated with a more methodical approach to assessing the problem and finding the solution (Schwarz & Skurnik, 2003). Our cognitive disequilibrium framework regards moods as secondary mediator variables in complex learning.

We believe that the cognitive disequilibrium framework goes a long way in explaining the emotions students experience during complex learning. This theoretical framework makes a number of predictions about the affective experiences during learning. Some of its predictions
have been tested in our analysis of emotion-learning connections in the context of intelligent tutoring systems and other advanced learning environments. These investigations are covered in the remainder of this chapter.

2. Learning Materials and Tasks

This section describes the learning materials and tasks that were used in our investigations of emotions and complex learning. Most of the learning environments have been computerized intelligent tutoring systems with pedagogical agents that hold natural language conversations with the student (Craig, Graesser, Sullins, & Gholson, 2004; D’Mello, Craig, & Graesser, 2009; D’Mello & Graesser, 2010, 2012, in press; Graesser, D’Mello, Chipma, King, & McDaniel, 2007; Graesser, Jackson, & McDaniel, 2007; Pour, Hussein, AlZoubi, D’Mello, & Calvo, 2010). The use of these pedagogical agents is appropriate for complex learning tasks that benefit from or require collaborative social interaction with expert tutors or instructors. The advantage of the computerized agents over human tutors lies in the consistency of the mechanisms that interpret student contributions and that strategically generate actions to scaffold learning. We have also collected data on learning environments without agents, such as problem solving and games (Baker, D’Mello, Rodrigo, & Graesser, 2010), preparation for a law exam (D’Mello, Lehman, & Person, in press), comprehension of illustrated texts (Strain & D’Mello, 2011), and argumentative writing. However, most of our work has been with intelligent agents in learning environments, particularly that of AutoTutor.

2.1 Learning Environments with Pedagogical Agents

Prior to the industrial revolution the typical way for students to learn a skill or subject matter followed an apprenticeship model that involved one-on-one conversations with a mentor, master, tutor, or instructor (Collins & Halverson, 2009; Graesser, D’Mello, & Cade, 2011;
The student and pedagogical expert would collaboratively work on tasks and problems as the student achieved new levels of mastery. The expert attended to the emotions of the student in addition to the student’s behavior and cognitive states. Available research on human tutoring supports the value of learning by collaborative social interaction (Graesser, D’Mello, & Cade, 2012; Graesser, Person, & Magliano, 1995). Learning gains are approximately 0.4 sigma for typical unskilled tutors in the school systems, when compared to classroom controls and other suitable controls (Cohen, Kulik, & Kulik, 1982), and vary from .2 to 2.0 for accomplished human tutors (Chi, Roy, & Hausmann, 2008; Roscoe & Chi, 2007; VanLehn et al., 2007).

Collaborative peer tutoring shows an effect size advantage of 0.2 to 0.9 sigma (Topping, 1996).

Pedagogical agents have recently been developed to serve as substitutes for human pedagogical experts. Some of these pedagogical agents express themselves with speech, facial expression, gesture, posture, and other embodied actions (Atkinson, 2002; Baylor & Kim, 2005; Biswas, Leelawong, Schwartz, Vye, 2005; Graesser, Jeon, & Dufty, 2008; Graesser, Lu et al., 2004; Gratch et al., 2001; Johnson & Valente, 2008; McNamara, O’Reilly, Rowe, Boonthum, & Levinstein, 2007; Millis et al., in press; Moreno & Mayer, 2004). The students communicate with the agents through speech, keyboard, gesture, touch panel screen, or conventional input channels. The agents help students learn by either modeling good behaviour and strategies or by interacting with the students in a manner that intelligently adapts to the students’ contributions. The agents may take on different roles: mentors, tutors, peers, players in multiparty games, or avatars in the virtual worlds. Single agents model people with different knowledge, strategies, personalities, physical features, and styles. Groups of agents model social interaction. Collectively these systems help students learn a variety of subject matters and skills, such as computer literacy,
electronics, physics, circulatory systems, critical thinking about science, foreign language, cultural practices, and reading and writing strategies.

Most of our work on emotions and complex learning has been conducted with AutoTutor, a pedagogical agent that helps students learn about computer literacy by holding a conversation in natural language. More recent work has been in a learning environment called Operation ARIES!, where students learn scientific reasoning by interacting with two agents in a conversational “trialog.” One agent is a tutor agent and the other a student peer agent. We now turn to a description of these two learning environments.

2.1.1 AutoTutor

AutoTutor is an intelligent tutoring system (ITS) that helps students learn topics in Newtonian physics, computer literacy, and critical thinking through a mixed-initiative conversational dialog between the student and the tutor (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Jeon, & Dufty, 2008; Graesser, Lu et al., 2004; Graesser, Wiemer-Hastings et al., 1999; VanLehn et al., 2007). AutoTutor’s dialogues are organized around difficult questions and problems (called main questions) that require reasoning and explanations in the answers. AutoTutor actively monitors the students’ knowledge states and engages them in a multi-turn conversational dialogue as they attempt to answer these questions. It adaptively manages the tutorial dialogue by providing feedback (e.g. “good job”, “not quite”), pumping the learner for more information (e.g. “What else”), giving hints (e.g. “What about X”), generating prompts to elicit specific words, correcting misconceptions, answering questions, and summarizing answers. The conversational moves of AutoTutor are guided by constructivist theories of pedagogy and scaffolds students to actively generate answers rather than by merely instructing students with well organized information delivery.
Figure 2 presents the interface of one version of AutoTutor along with a short dialogue with a student. The interface includes a main question, an animated conversational agent, an auxiliary diagram, a window with the dialogue history, and a window for the student response in a single conversational turn. A multi-sentence explanation is needed to answer the main question, “When you first turn on a computer, how is the operating system first activated and loaded into RAM?” However, students never give a lengthy complete answer in a single turn. Instead, the conversation is distributed over many turns with over 90% of the students’ turns ranging from one word to two sentences. The conversation excerpt in Figure 2b shows some of the dialogue moves that AutoTutor generates to move the conversation along and cover all of the sentences in a complete answer to the main question. There is a pump, hint, feedback, and an assertion. In most versions of AutoTutor, the student types in the response for each conversational turn, but some versions can accommodate spoken student contributions with speech-to-text recognition (D’Mello, Dowell, & Graesser, in press; D'Mello, King, & Graesser, 2010).

It is beyond the scope of this chapter to describe the mechanisms of AutoTutor in detail because our focus is on emotions and the mechanisms of AutoTutor have been described in other publications cited throughout this chapter. However, some points will be made about the scaffolding mechanism, dialogue moves, and learning gains.

Scaffolding mechanisms. As mentioned, AutoTutor’s dialogues are organized around difficult questions and problems that require reasoning and explanations in the answers. For example, the main question (“When you turn on the computer, how is the operating system first activated and loaded into RAM?”) calls for approximately 5 sentences in an ideal answer.
AutoTutor tries to get the student to articulate the content of the 5 sentences rather than telling the student the answer. In other words, AutoTutor elicits rather than lectures. In order to track what the student contributes in the distributed conversation, AutoTutor performs *pattern matching* computations between the student’s accumulation of verbal content and the 5 sentences in the ideal answer. The pattern matching algorithms include advances in computational linguistics and information science, such as latent semantic analysis (Landauer, McNamara, Dennis, & Kintsch, 2007), regular expressions (Jurafsky & Martin, 2008), and overlap in content words weighted by word frequency. There are invariably missing words and ideas in the student’s accumulating contributions, so AutoTutor generates dialogue moves to achieve *pattern completion* and fill out all of the content of the ideal answer. These dialogue moves include pumps, hints, and question prompts for specific words to be filled in. If all else fails, AutoTutor generates assertions to fulfill pattern completion and finish answering the main question.

**Dialogue moves.** Some of the dialogue moves would be expected to have a noticeable impact on student emotions. What might these be?

1. **Feedback.** AutoTutor’s positive feedback should elicit a positive emotional valence in the student whereas negative feedback should elicit a negative emotional valence.

2. **Hints.** AutoTutor’s hints may not be fully understood or appreciated, which runs the risk of eliciting confusion or frustration in the student.

3. **Corrections.** AutoTutor sometimes corrects student errors, which is likely to elicit a negative valence or confusion in the student.

4. **Main question.** The questions are difficult, so students are likely to experience a number of emotions associated with cognitive disequilibrium.
When the student is on the roll and expressing a stream of correct answers, then they would be expected to experience a positive emotional valence, such as flow/engagement. When the conversation loses coherence and connections to what the student is thinking, the student would be expected to experience a negative emotional valence, such as confusion or frustration. In the rare cases when the student has to listen to a lengthy lecture or extremely dense material from AutoTutor, there is the risk of the student experiencing boredom and disengagement. These are some of the a priori expectations of how learning and emotions may be related, but once again there is precious little research on moment-to-moment learning-emotion relations.

**Learning gains.** The learning gains of AutoTutor have been evaluated in over 20 experiments conducted during the last 15 years. Assessments of AutoTutor on learning gains have shown effect sizes of approximately 0.8 standard deviation units in the areas of computer literacy (Graesser, Lu et al., 2004) and Newtonian physics (VanLehn, Graesser et al., 2007) compared with reading a textbook for an equivalent amount of time. The assessments of learning gains from AutoTutor have varied between 0 and 2.1 sigma (a mean of 0.8), depending on the learning performance measure, the comparison condition, the subject matter, and the version of AutoTutor. Measures of learning in these assessments have included: (1) multiple choice questions on shallow knowledge that tap definitions, facts and properties of concepts, (2) multiple choice questions on deep knowledge that taps causal reasoning, justifications of claims, and functional underpinnings of procedures, (3) essay quality when students attempt to answer challenging problems, (4) a cloze task that has subjects fill in missing words of texts that articulate explanatory reasoning on the subject matter, and (5) performance on problems that require problem solving. Assessments of learning in various conditions with various measures have uncovered the following generalizations.
(1) **AutoTutor versus reading a textbook.** Learning gains with AutoTutor are superior to reading from a textbook on the same topics for an equivalent amount of time.

(2) **AutoTutor versus expert human tutors.** Learning gains of AutoTutor are nearly the same as the gains of accomplished human tutors via computer mediated communication.

(3) **Deep versus shallow tests of knowledge.** The largest learning gains from AutoTutor have been on deep reasoning measures rather than measures of shallow knowledge.

(4) **Zone of proximate development.** AutoTutor is most effective when there is an intermediate gap between the learner’s prior knowledge and the ideal answers of AutoTutor. AutoTutor is not particularly effective in facilitating learning in students with high domain knowledge and when the material is too much over the learner’s head.

### 2.1.2 Operation ARIES!

**ARIES (Acquiring Research Investigative and Evaluative Skills).** This system was developed in a research collaboration between University of Memphis, Northern Illinois University, and Claremont Graduate School (Millis et al., in press). ARIES teaches scientific critical thinking through two or more animated pedagogical agents. One agent in ARIES, called the tutor-agent, is an expert on scientific inquiry. The other agent is a peer of the human student. There are agents that take on other roles, but these will not be addressed in the present context. A 3-way conversation transpires between the human student, the tutor agent, and the student agent. The human students interact with both agents by holding mixed-initiative “trialogs” in natural language. ARIES also is a serious game with a story narrative, interactive text, testing modules, and cases in which the student critiques studies on scientific methods.

It is the case study modules and associated trialogs that are of direct relevance to this chapter. A series of cases are presented to the student that describes experiments which may or
may not have a number of flaws with respect to scientific methodology. For example, a case study may describe a new pill that purportedly helps people lose weight, but the sample size is small and there is no control group. The goal is to identify the flaws and express them in natural language. Some studies had subtle flaws while others were flawless; this made the flaw detection task quit challenging.

3. Emotions that Occur during Difficult Learning Materials and Tasks

This section describes the studies we conducted to explore the moment-to-moment emotions that occur while college studies learn technical material with AutoTutor and ARIES. At the risk of killing the suspense, we found that the primary learning-centered emotions were confusion, frustration, boredom, flow/engagement, delight and surprise. Anxiety surfaced when there are high stakes, as in the case of an examination. Curiosity was also experienced when there was freedom of choice or when intrinsic motivation was high. These emotions emerged as being important after conducting several studies on a large number of learning and problem solving tasks (Baker et al., 2010; D’Mello & Graesser, 2012, in press). Most of the initial experiments were conducted with AutoTutor, whereas our recent work on ARIES focused on the role of confusion during learning. In addition to identifying the distribution of emotions during learning, we report studies on the duration of emotions (emotion chronometry) and transitions between emotions (emotion dynamics). There are also some studies that examine what emotions correlate with learning gains and, in the case of confusion, whether confusion might cause or mediate an increase in learning.

3.1 What Emotions Occur During Complex Learning?

We have implemented several methods to track the moment-to-moment emotions that occur during learning. Most of these methods were non-invasive, meaning they did not disrupt
the normal stream of learning by asking students what their emotions were and thereby biasing the course of their learning. But some methods did explicitly poll the students on their emotions during learning by collecting self-report measures. We also have tracked the emotions with computer software that analyzes language, speech, facial expressions, and body movements. In all of these studies, the goal is to measure affective states at many points in time during the learning sessions that typically last 30 minutes to 1 hour.

3.1.1 Trained Observers

Our first study simply had trained judges observe college students interacting with AutoTutor (Craig, Graesser, Sullins, & Gholson, 2004) and observe the emotions that occur. These emotions were also correlated with the learning gains from AutoTutor on the subject matter of computer literacy. Five trained judges observed six different affect states (confusion, frustration, boredom, flow/engagement, eureka, and neutral). The participants were 34 college students who had low knowledge of computer literacy; they answered 10 or fewer questions out of 24 multiple choice questions on a pretest. Expert 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. The relative frequency of these six emotions were correlated with proportional learning gains, defined as: \[
\frac{(\text{posttest scores} - \text{pretest scores})}{(1.0 - \text{pretest scores})}.
\]

This observational study revealed that the most frequent emotions that occurred during complex learning of computer literacy with AutoTutor were flow/engagement (45%), boredom (18%), and confusion (7%), with frustration and eureka being quite rare. There was only one recorded eureka experience in the over 10 hours of tutoring among the 34 students.
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, but that would no doubt be explained by the low frequency of these emotions. Follow-up research on AutoTutor (D’Mello & Graesser, 2012, in press; Graesser, D’Mello, Chipman, King, McDaniel, 2007) revealed that confusion is the best predictor of learning gains among the various emotions investigated. The positive correlation between confusion and learning is consistent with a model that assumes that cognitive disequilibrium is an important correlate with deep learning, as discussed earlier. The findings that learning correlates negatively with boredom and positively with flow/engagement are consistent with predictions from Csikszentmihalyi’s (1990) analysis of flow experiences.

3.1.2 Emote Aloud Protocols

An emote-aloud procedure collects spoken verbal expressions of emotions while the students complete a task, in this case learning with AutoTutor (Craig, D’Mello, Witherspoon, & Graesser, 2008; D’Mello, Craig, Sullins, & Graesser, 2008). The emote aloud procedure is analogous to the traditional think aloud procedure (Ericsson & Simon, 1993) except that the students are instructed to articulate their emotions instead of the cognitive content that typically surfaces in think aloud protocols. Pilot studies revealed that most students do not know what it means to express emotions so they need some guidance on what the alternative emotions might be and how to label them. Therefore, we listed and defined a set of emotions that they might be experiencing while learning from AutoTutor: Confusion, frustration, boredom, anger, contempt, curiosity, disgust, and delight/eureka. These affective states were defined before the students began the 90-minute tutoring session. Flow/engagement was not included in this study because
of the concern that asking people to report on their flow experiences would disrupt those experiences. The students also had the freedom to express other emotions that they were experiencing. Audio of these *emotes* were recorded and transcribed for analysis.

We collected emote aloud protocols from only a small sample of students (N=7) because of the challenges of transcribing the data and linking them to events in AutoTutor. Nevertheless, the data were reasonably informative. First, there were substantial differences among students with respect to expressing emotions in the emote-aloud task. The mean number of emotes during the 90-minute session was 31, ranging from 6 to 89. The emote fluency was extremely low for some and high for others so the emote aloud procedure is best reserved for the more expressive individuals. Second, the percentages of emotions revealed that the most prevalent emotions were frustration (26%), confusion (25%), boredom (20%), and delight (14%), whereas the remaining emotions comprised only 10% of the observations. There were very few emotes that were not on the list of emotions defined for the students. Although delight/eureka was relatively well reported, we suspect that this response functionally signified delight from giving a correct answer rather than a deep eureka experience.

### 3.1.3 Identification of Emotions by Learners, Peers, Trained Judgers and Expert Teachers

The observational and emote-aloud studies collected emotion labels while students concurrently learned from AutoTutor. The studies reported in this subsection used an off-line retrospective emotion judgment protocol to poll the students’ emotions (D’Mello, Craig, & Graesser, 2009; Graesser, D’Mello et al., 2008). College students interacted with the AutoTutor system for 32 minutes without any interruptions of the normal learning process. We recorded videos of their faces, their posture while sitting down, and the computer screens during their interactions with AutoTutor. The facial expressions and computer screen views were integrated
into a single video for subsequent collection of retrospective emotion judgments. The screen capture included the tutor’s synthesized speech, printed text, students’ responses, dialogue history, and images, thereby providing the context of the tutorial interaction.

Retrospective emotion judgments were provided by the student learners themselves (Self), untrained peers (Peers), and two trained researchers (Experts) with considerable experience interacting with AutoTutor and with the Facial Action Coding System developed by Ekman (Ekman & Friesen, 1978), We also collected these affect judgments from two experienced high school teachers (D’Mello, Taylor, Davidson, & Graesser, 2008). Therefore, the same tutorial video was analyzed by multiple judges with very different training on emotions and pedagogy.

There was a systematic procedure for collecting retrospective emotion judgments. After the learner was finished interacting with AutoTutor, the participant viewed the videos and gave judgments on their emotions at 20 second intervals. The video automatically stopped at these points, called mandatory observations. They judged whether any of 7 emotions had occurred: confusion, frustration, boredom, flow/engagement, delight/eureka, surprise, and neutral. There was a checklist of emotions for them to mark, along with an “other” category for them to provide additional emotions that they viewed as relevant. They were also instructed to indicate any affective states that were present in between the 20-second stops (called voluntary observations). If the student was experiencing more than one affective state, judges were instructed to mark each state and indicate which was most salient. Our sample of observations had over 2500 mandatory judgments and 1000 voluntary judgments when considering the sample of 28 college students.
Judgments on the emotions were also collected from judges other than the student learner (Self). After the student learners were finished making their judgments on affective states, they served as peer judges a week later by making judgments on another student’s emotions during their AutoTutor interaction. Two trained expert judges also judged each participant’s emotions during AutoTutor interaction.

We examined the percentages of judgments that were made for each of the emotion categories, averaging over the 4 judges. The most common affective state was neutral (37%), followed by confusion (21%), flow/engagement (19%), and boredom (17%); the remaining states of delight, frustration and surprise totaled 7% of the observations. The voluntary emotion judgments were expected to include more visible and salient emotions (with theoretically higher physiological arousal) compared to the more subtle emotions at the mandatory 20-second intervals. The more salient voluntary points had a rather different distribution of emotions. The most prominent emotion was confusion (38%), followed by delight (19%) and frustration (19%), whereas the remaining affective states comprised 24% of the observations (boredom, surprise, flow, and neutral, in descending order). Most of the time the students were either in a neutral state or were experiencing a subtle emotion (boredom or flow/engagement).

When these data are considered in conjunction with the observational data and emote aloud data reported earlier, the predominate emotions during learning are confusion, frustration, boredom, and flow/engagement, with delight and surprise occasionally occurring with lower relative frequencies. These are the emotions that we call learner-centered emotions. They are very different from Ekman’s big six emotions of happiness, sadness, fear, anger, disgust, and surprise.
It is very difficult to establish the ground truth in declaring what the emotions the student is actually experiencing. There is no reason to believe that the student is the most knowledgeable judge, as every clinical psychologist would tell us. Peers would unlikely be the most valid judges because students are not trained in school on the fundamentals of human emotion and how to recognize emotions of others. The judgments of trained experts are presumably the most valid, but how does one know without any defensible gold standard? An aggregate score that considers the different viewpoints may be the best proxy for a gold standard.

Given there is no ideal gold standard, it is important to examine the level of agreement among the different judges. To what extent is there was agreement in the emotion judgments provided by the Self, Peer, Expert1 and Expert2? The design of this study allowed us to assess the reliability of judgments by computing Cohen's kappa scores between 6 pairs of judges: Self-Peer, Self-Expert1, Self-Expert2, Peer-Expert1, Peer-Expert2, and Expert1-Expert2. Cohen’s kappa scores served as the metric of inter-judge agreement because it adjusts for base rate levels and a quantitative scale that varies from 0 (chance) to 1 (perfect agreement). Table 1 shows mean kappa scores as a function of the six combinations of judges, with separate columns for mandatory, voluntary, and all observations. The scores in Table 1 revealed that the two Experts had the highest agreement, the Self-Peer pair had near zero agreement, and the other pairs of judges were in between. An ANOVA was performed on the left column of scores that included all observations, namely the mandatory plus voluntary observations. The results confirmed that there were significant differences in kappa scores among the six pairs, $F(5, 135) = 33.34$, $MSe = .008$, $p < .01$. Post hoc tests revealed that the Self-Peer pair had the lowest inter-judge reliability scores ($p < .05$) when compared to the other five pairs and the two Experts had significantly higher kappa scores than the other five pairs. The same pattern of kappa scores
occurred when the mandatory and the voluntary observations were analyzed separately. Differences among pairs were quite pronounced for the voluntary judgments; the Expert pair achieved a kappa score as high as 0.71 in contrast to a very low 0.12 kappa for the Self-Peer pair.

These findings on inter-judge kappa scores support a number of conclusions. First, the agreement scores are quite modest so there is a fundamental challenge on establishing the ground truth of the learner-centered emotions. Second, the agreement scores are considerably higher (more than double) for the more salient voluntary observations than the mandatory observations. Third, training on Ekman’s facial action coding system and tutorial dialogue can enhance the reliability and accuracy of judgments of affective states. Indeed the agreement between experts on voluntary observations reached a respectable 0.71 kappa. Fourth, peers are not good judges of the emotions of students, with kappa scores drifting toward 0.

A follow-up study revealed that accomplished master teachers are similarly not adept at recognizing the emotions of the students (D’Mello, Taylor, Davidson, & Graesser, 2008). Their kappa scores showed patterns similar to Peers in deviating from the Self and Experts. Contrary to suggestions by Lepper (Lepper & Woolverton, 2002), accomplished teachers did not seem to be very adept at detecting the learners’ emotions. Untrained peers and accomplished teachers do not seem to be very proficient at judging the emotions of the learner. Coders need to be trained on emotion detection before respectable agreement scores emerge (Ekman, Sullivan, Frank, 1999). Once again, however, there is no ground truth on what the actual emotions are.

There was also a follow up study conducted by D’Mello, King, Entezari, Chipman, and Graesser (2008) that replicated the above analyses but with the additional channel of speech recognition. That is, it was a replication of the multiple-judge study with the exception that 30
students spoke their responses to a speech-enabled version of AutoTutor. There was a retrospective emotion judgment procedure with judgments provided by the Self and Peer.

3.1.4 Comparisons of Different Computer-Based Learning Environments.

Baker, D’Mello, Rodrigo, and Graesser (2010) tracked the emotions in three different computerized learning environments in order to assess the generality of our claims about the prevalence of learning-centered emotions. The first environment was AutoTutor, as we have already reported. The second involved students interacting with the Aplusix II Algebra Learning Assistant (Nicaud, Bouhineau, Mezerette, Andre, 2007). The third was The Incredible Machine: Even More Contraptions (Sierra Online Inc., 2001), a simulation environment in which students complete a series of logical puzzles. Together these three environments included different populations (Philippines versus USA, high school students versus college students), different methods (quantitative field observation versus retrospective self-report), and different types of learning environment (dialogue tutor, problem-solving game, versus ITS with problem-solving). Baker et al. (2010) investigated the following affective states: confusion, frustration, boredom, flow/engagement, delight, surprise, and neutral. Learning gains or performance was also measured in these environments.

We analyzed the relative prevalence of different emotions in the three environments. Boredom was frequent in all learning environments, was associated with poorer learning, and was associated with the dysfunctional behavior called gaming the system (i.e., mechanically using system facilities to trick the system into providing answers rather than learning the domain knowledge). Frustration was considerably less frequent, less associated with poorer learning, and was not an antecedent to gaming the system. Confusion was consistently observed in all learning environments, whereas there were informative differences in the occurrence of flow/engagement.
Experiences of delight and surprise were rare. Baker et al. (2010) advocated that significant effort should be put into detecting and productively responding to boredom, frustration, and confusion. There should be a special emphasis on developing pedagogical interventions to disrupt the “vicious cycles” which occur when a student becomes bored and remains bored for long periods of time to the point of disengagement or frustration (D’Mello & Graesser, in press; D’Mello, Taylor, & Graesser, 2007). More will be said about sequences of emotions later in this chapter.

In another study conducted by D’Mello, Lehman, and Person (in press) 41 students preparing for a law school entrance examination solved 28 difficult analytical reasoning problems from the law school admissions test (LSAT). Their facial expressions were recorded in addition to the computer screen. Students later completed a retrospective emotion judgment procedure. They made affect judgments at pre-specified points when the videos automatically paused. These affect judgments points were: (1) a few seconds after a new problem was displayed, (2) halfway between the presentation of the problem and the submission of the response, and (3) three seconds after the feedback was provided. In addition to these three pre-specified points, students were able to manually pause the videos and provide affect judgments at any time (voluntary judgments). Students judged their emotions from the following alternatives: confusion, frustration, boredom, flow, contempt, curiosity, eureka, anger, disgust, fear, happiness, sadness, surprise, and neutral. The results revealed that boredom, confusion, frustration, curiosity, and happiness (e.g., delight) were the major emotions that students experienced during problem solving, whereas anxiety was another important emotion. The emotion of anxiety is expected to surface more frequently when students anticipate evaluation and high stakes tests.
3.1.5 Automated Detection of Emotions

Another method of classifying emotions during learning is through computerized methods. Automated emotion detection has been a central priority in our research program because one of our goals was to develop an emotion-sensitive AutoTutor that tracks the student’s emotions automatically and generates dialogue moves designed to optimize learning in a manner that is responsive to the student’s emotional and cognitive states (D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; D’Mello & Graesser, 2010, in press; D’Mello, Picard, & Graesser, 2007; Graesser, Jackson, & McDaniel, 2007). It is beyond the scope of this chapter to describe the computational mechanisms of our automated emotion detectors, but we will give some highlights of the channels investigated and how the automated mechanisms compare with human judgments.

Most of our work on automated emotion detection has concentrated on three channels. These channels include the discourse interaction history (D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; Graesser, D’Mello et al., 2008), facial actions (McDaniel et al., 2007), body movements (D’Mello, Dale, & Graesser, in press; D’Mello & Graesser, 2009), and combinations of these three channels (D’Mello & Graesser, 2010, 2012, in press). Figure 3 depicts these three channels in the context of the AutoTutor system. The discourse interaction history includes events stored in the AutoTutor log file, the speech acts of student and tutor turns, and the knowledge states achieved by the student during the tutorial dialogue. An analysis of the discourse interaction history provides a model of the context of an emotional expression. The facial actions and expressions are tracked by different systems (FaceSense, Mindreader) developed in Picard’s Affective Computing Laboratory (el Kaliouby & Robinson, 2005), who collaborated on the project. A body posture pressure measurement system manufactured by
Tekscan tracks motions of the body against the seat and back of the chair. In addition to these channels we have also investigated a haptic pressure sensor for the mouse (supplied by MIT), a keyboard pressure sensor, and acoustic-prosodic features obtained from students who gave spoken contributions to AutoTutor through the Dragon speech recognition system.

Insert Figure 3 about here

How well did the automated emotion detectors fare in detecting emotions? We compared the computer’s prediction with each of the judge’s decisions on a sample of observations collected in the AutoTutor study reported in 3.1.3. The results of these decisions were moderately encouraging, resulting in reliability scores on par with the novice judges but not the experts. The major accuracy metric we adopted in most of the analyses was a score on a binary scale that had a 50% chance of being in some emotion category E compared with a neutral state. It is beyond the scope of this chapter to present details on the performance results, which varied among channels, combinations of channels, and emotion categories. Instead we give some highlights of results for each channel and cues that predict emotion detection.

**Dialogue Interaction history.** Our results of the best predictive model indicated showed accuracies of 63%, 77%, 64%, 70% and 74%, in discriminating confusion, frustration, boredom, flow, and delight from neutral. The average across emotions was 70%. If we were to transform these scores to values comparable to kappa scores [i.e., $2 \times (\text{score} - 0.5)$], the quantities would be .26, .54, .28, .40, and .48, respectively, or .39 overall. Such kappa scores are comparable to accuracy scores reported by other researchers in the literature who have attempted automated emotion detecting systems.

The dialogue cues that trigger the emotions are quite different for the different emotions. The cues that accompany confusion tend to be short student responses, frozen student expressions (such as “I don’t know,” “Uh huh”), speech acts by the tutor that are indirect (such
as hints), and early time phases during the student’s initial attempts to solve the problem or answer the questions posed by the tutor. In contrast, the cues that accompany frustration are negative tutor feedback and student responses that are locally good ideas but not globally good. Flow/engagement tends to occur with lengthier answers, early phases of the dialogue, and after positive tutor feedback. Boredom tends to occur in later phases in the session or particular problem and when the tutor tends to lecture with direct assertions.

**Facial Expressions.** The fully automated facial expression analysis system was in development at the time of this evaluation, so we relied on human annotated facial features to infer relationships between facial movements and emotions. We adopted the Ekman and Friesen (1978) Facial Action Coding system in an analysis of facial action units (AUs). This system specifies how emotions can be identified on the basis of 58 facial behaviors and the muscles that produce them. The kappa scores between trained human judges in identifying the AUs in the faces reached a respectable level of agreement (0.72). These trained judges identified the AUs for a sample of emotions displayed during the AutoTutor sessions reported in section 3.1.3. The numbers of facial expressions in the sample for confusion, frustration, boredom, delight, and neutral were 59, 47, 26, 43, and 37, respectively. There were not enough surprise emotions in the sample for an analysis. There were sufficient observations with flow/engagement, but these were not included in the sample because we concluded at that time that flow/engagement was not substantially different from neutral in facial depictions.

The results of the best predictive model showed accuracies of 76%, 74%, 60%, 60% and 90%, in discriminating confusion, frustration, boredom, flow, and delight from neutral. The average across emotions was 72%. Transformation of these scores to values comparable to kappa scores showed quantities of .52, .48, .20, .20, and .80, respectively, or .44 overall. The
classifiers were more successful in detecting emotions that are manifested with highly animated facial activity, such as delight and confusion, than emotions that are more subtly expressed (boredom, flow).

The facial cues were quite different for the different emotions, as would be expected. Examples of these emotions analyzed, excluding neutral, are presented in Figure 4. An AU was considered distinctive for an emotion if its presence significantly differed from a neutral expression. We found that confusion was manifested by a lowered brow (AU 4), the tightening of the eye lids (AU 7), and a notable lack of a lip corner puller (AU 12), as depicted in Figure 4a. Several AUs were associated with delight, which is depicted in Figure 4b. There was presence of the lid tightener (AU 7), lip corner puller (AU 12), lips part (AU25), and jaw drop (AU26) coupled with an absence of the eye blink (AU 45). These patterns are quite similar to a smile. Boredom (see Figure 4c) was not easily distinguishable from neutral on the basis of the facial features. Boredom resembles an expressionless face. Frustration is a state that is typically associated with significant physiological arousal, yet the facial features we tracked were not very good at distinguishing this emotion from neutral (see Figure 4d). The only significant correlation with frustration was obtained for the lip corner puller (AU 12), perhaps indicative of a half smile with an affinity to disgust. Students apparently tend to disguise frustration because it is not socially appropriate in most contexts (Hoque, Morency, & Picard, 2011).

Insert Figure 4 about here

It is apparent from these analyses that facial expressions have distinctive signatures for some emotions but not others. Confusion, delight, and presumably surprise have obvious facial patterns whereas it is difficult to distinguish boredom, frustration, and flow/engagement from neutral. These latter emotions need to be identified by communication channels other than the
face. For example, posture and the dialogue interaction history differentiates these latter emotions. The fact that some emotions are not manifested in the face of course limits the accuracy of computer automated detection of emotions. To complicate matters, our current automated facial detector of confusion, delight, and surprise is reasonably accurate for some students, but not for other students. Some students would require high resolution technologies to handle subtle discriminations in facial movements, skin texture, and contrasts between the skin versus the brow, eyes, and lips.

**Body Posture.** The results of the best predictive model showed accuracies of 65%, 72%, 70%, and 74% in discriminating confusion, frustration, boredom, and flow/engagement from neutral. The average across emotions was 70%, the same level of accuracy as dialogue interaction history and facial expressions. Transformation of these scores to values comparable to kappa scores showed quantities .30, .44, .40, and .48, respectively, or .41 overall. These results confirm that posture can be a viable channel in inferring students’ emotions, particularly those of frustration, boredom, and flow/engagement, the very same emotions for which facial expressions were not particularly diagnostic.

The models that detect emotions from body posture range from pressure features of the body against the chair to complex dynamical systems models that track fine-grained changes in pressure over time (D’Mello, Dale, & Graesser, in press; D’Mello & Graesser, 2009). It is beyond the scope of this chapter to specify these models but some highlights will be given. One simple method to track body movements consists of analyzing the individual frames of the two posture pressure maps placed on the back and the seat of the chair that the student is seated on. The *current frame* refers to the pressure patterns on the back and the seat at the time of the affective experience of interest. We compute the average net force against the seat and back
during the current time frame and also compare that average to the readings 2 seconds before and 2 seconds after the current change, which is an indicator of the change in pressure patterns.

There are also other measures inspired by dynamical systems theories that analyze the fine-grained fluctuations over time, with shifts in the noise patterns (called white, pink, and brown noise).

The different emotions showed different patterns of posture features. During episodes of boredom, the learners leaned back while they leaned forward when they were engaged. They also assumed an attentive posture when they were confused or frustrated. Importantly, there was an increase in the rate of change of fluctuations in body movements when there was an emotion compared to the neutral state. One pattern that emerged from our dynamical systems analysis indicated that those emotions that accompany cognitive disequilibrium (e.g., confusion and frustration) exhibit characteristics correlated with a more systematic pink noise initially but then shifts to an unstructured white noise (D’Mello, Dale, & Graesser, in press). The dynamics of the body apparently play an important role in differentiating frustration, boredom, flow/engagement, and neutral states, which are not saliently manifested in facial expressions.

**Combinations of channels.** It is likely of course that emotion detection is best predicted by a combination of the dialogue, face, and body characteristics. D’Mello and Graesser (2010) explored a number of computational models that combined these channels in predicting emotions.

The previous classification analyses indicated that channels differ in being diagnostic of particular emotions. The posture sensor would be the sensor of choice for affective states that do not generate overly expressive facial expressions, such as boredom and flow/engagement. On the other hand, the affective states of confusion, delight, and surprise, which are accompanied by
significant arousal, are best detected by monitoring facial features. The negative affective state of frustration is typically disguised and therefore difficult to detect by the face and body, but the dialogue features come the rescue in its detection. Taken together, detection accuracies were over 77% (roughly a .55 kappa) when particular emotions are aligned with the optimal sensor channels. So one way of detecting the emotion is to select the decision of the channel that has the highest resonance with a particular emotion. This approach would be consistent with the classical Pandemonium model (Franklin, 1995; Selfridge, 1959).

A feature-level sensory fusion model takes a different approach to combining channels (D’Mello & Graesser, 2010). Fusion at the feature level involves grouping features from the various sensors before attempting to classify emotions. One might expect that the combination of features from 2 or 3 sensors is higher than either channel alone. Superadditivity would occur if there are improvements in multisensory fusion over and above the maximum unisensory response or an additive combination of the contributions from the different sensors. Redundancy would occur if multisensory fusion equals the maximum unisensory response. We discovered that redundancy among channels was much more prevalent than superadditivity or an additive combination of sensor contributions. However there was a modest significant improvement of multisensory fusion over the maximum single channel.

We evaluated the accuracy of discriminating between boredom, confusion, frustration, and neutral with a split-half evaluation method. Fourteen of the 28 students were randomly selected and their instances were assigned to the training set. Instances from the remaining 14 students were assigned to the test set. Discriminant models were constructed from the training instances and evaluated on the testing instances. The discriminant models yielded a 48.8% accuracy on the unseen test set for discriminating between boredom, flow/engagement,
confusion, frustration, and neutral. These results are positive because they imply that these moderate accuracy scores can be expected in real-world situations where the affect detector has to classify the emotions of unknown students.

3.2 Temporal Dynamics of Emotions

The affective experiences that accompany learning are transient and dynamically change during learning rather than being persistent and static. The emotions evolve, decay, and change throughout the learning experience as the student interacts with the complex learning environment. In order to get a better understanding of these dynamics, we have conducted analyses on the duration of emotions (D’Mello & Graesser, 2011) and the transitions between emotions during learning with AutoTutor (D’Mello & Graesser, 2012, in press; D’Mello, Taylor, & Graesser, 2007).

3.2.1 Duration of Emotions

There is a gap in the scientific literature on the persistence of the learning-centered emotions so we conducted some research to explore how long the various emotions persist (D’Mello & Graesser, 2011). Stated differently, what is the half-life of emotions after the emotions begin? The expected temporal chronometry would specify a point in time that an emotion is started, a duration from the start-point to the peak of the emotion, a duration of emotional experience around the peak, and a decay or dampening of the emotion until base level is achieved. The dampening is expected to follow an exponentially decreasing function, like most extinction curves. A simple metric of the half-life is the duration from the start-point of an emotion and the point in time at the 50% relative distance between the peak and base level.

Our database included the study on AutoTutor reported in section 3.1.3, where the emotions were polled every 20 seconds (mandatory observations) and the judges could identify
emotions in between these set points (voluntary observations). The polling of 20 second increments is no doubt crude for tracking the duration of emotions because some last only 2 seconds or less (Ekman, 1992). Nevertheless, we could collect some information on the relative durations of the different learning-centered emotions.

There is some foundation for predicting the following relative ordering in half-life durations: (Delight = Surprise) < (Confusion = Frustration) < (Boredom = Engagement/Flow). The foundation appeals to the hierarchy of goals that guide organized behavior, goal achievement, and the interruption or blockage of goals (Mandler, 1976; Ortony et al., 1988; Stein et al., 2008). The primary goals in learning tasks with AutoTutor are to comprehend the material and solve difficult reasoning tasks. The students are typically in a prolonged state of either (a) flow/engagement as they pursue the superordinate learning goal of handling the material or (b) disengagement (boredom) when they encounter a major persistent goal blockage and give up pursuit of the superordinate learning goal. Boredom and flow/engagement should have the longest duration under this theoretical analysis. In contrast, confusion and frustration occur when there is novel information, a discrepancy between the materials and world knowledge, and goals that are blocked. The students initiate a subgoal of assimilating the materials or resolving the impasse through effortful comprehension, reasoning, and problem solving. Confusion and frustration are affiliated with subgoals, so they should be shorter than the states of flow and boredom that address the major goal. In the case of extreme novelty or an unexpected outcome, the event evokes surprise, a short-lived emotion. When there is an event that triggers the achievement of a goal, the emotion is positive, such as delight or even one of those rare eureka experiences (Knoblich, Ohlsson, & Raney, 2001). Previous research on delight and surprise support the claim that these emotions are typically quite brief (Ekman, 1992; Rosenberg, 1998).
D’Mello and Graesser (2011) developed exponential decay functions for each of the emotions in their analysis of the half-life of emotions. The models supported a tripartite classification of learning-centered emotions: persistent emotions (boredom, flow/engagement, and confusion), an emotion of intermediate duration (frustration), and transitory emotions (delight and surprise). This pattern somewhat confirms the aforementioned predictions stemming from goal-appraisal theories of emotion, with the exception that confusion was categorized as a persistent rather than an intermediate emotion. The emotions expected to have intermediate durations will no doubt depend on the level of challenge and scaffolding in the learning environment.

**3.2.2 Transitions between Emotions**

Transitions from one emotion to another are influenced by the difficulty of the materials, the dialogue interaction between student and computer, the student’s level of mastery, and a host of other factors that were discussed in the context of the cognitive disequilibrium framework (see Figure 1 and section 1.3). One way to test or discover the moment-to-moment transitions in emotions is to document the emotion transitions in a transition matrix and to identify the events in the learning environment that explain these transitions. These analyzes have been conducted on the AutoTutor data set described in section 3.1.3 (D’Mello & Graesser, 2012, in press; D’Mello, Taylor, & Graesser, 2007).

A quick glimpse of emotion changes can be inspected by plotting the coded emotions as a function of time in the AutoTutor sessions. These emotions are plotted alphabetically by label. Figure 5 presents two such plots. The student in Figure 5a is remarkably stable over time in a neutral state but occasionally experiences confusion delight, and boredom. The student in Figure 5b is in an emotional rollercoaster vacillating between confusion and flow, with experiences of
all of the other emotions except for neutral. As discussed in section 1.3, this oscillation between confusion and flow is compatible with the cognitive disequilibrium framework.

We are particularly interested in the transition from one emotion to a different emotion in this analysis. The repetition of the same emotion is of course important and was captured in our analysis of emotion duration. However, it is the change in emotion category that is of interest in the present analysis. A metric is needed that computes the likelihood of shifting from emotion category at time $t$ to another emotion category at time $t + 1$ in a way that quantitatively adjusts for the base rate likelihood of the emotion category at $t + 1$.

The desired metric and transition analyses have been reported in D’Mello, Taylor, and Graesser (2007) and in D’Mello and Graesser (in press). The metric expressed in the equation below computes the relative likelihood of transitioning from an emotion at time $t$ to a subsequent emotion at time $t + 1$. This likelihood is represented as $L(M_t \rightarrow M_{t+1})$, where $M_t$ is the current emotion and $M_{t+1}$ is the next emotion. The denominator in the equation is simply a normalization factor.

$$L(M_t \rightarrow M_{t+1}) = \frac{\Pr(M_{t+1}|M_t) - \Pr(M_{t+1})}{1 - \Pr(M_{t+1})}$$

The metric was used to compute six data sets, one for each target emotion (confusion, frustration, boredom, flow, delight, and surprise). The metric permitted us to directly compare the relative likelihood that individuals in an affective state at time $t$, will change to another affective state at time $t + 1$. Repeated-measures ANOVAs, with the student as the unit of analyses, were then computed to determine if there were significant differences between the current emotion and the emotions that immediately followed. The major hypotheses of the model
were tested by performing time-series analyses on the data from the multiple-judge study in section 3.1.3 and the follow up speech recognition study with AutoTutor (D'Mello & Graesser, in press; D'Mello, Taylor, & Graesser, 2007).

Cognitive disequilibrium theory makes a number of predictions about the transitions between the learning-centered emotions (see Section 1.3 and Figure 1). Learners who are in a flow/engaged state will experience confusion when an impasse is detected. They engage in effortful problem solving activities in order to resolve the impasse and restore equilibrium. Equilibrium is restored when the impasse is resolved and learners revert back into the flow/engaged state. However, confusion transitions into frustration when the impasse cannot be resolved, the student gets stuck, and important goals are blocked. Furthermore, persistent frustration may transition into boredom, a crucial point at which the learner disengages from the learning process.

The results confirmed the presence of confusion--flow/engagement and boredom--frustration oscillations as well as confusion to frustration transitions (see Section 1.3 and Error! Reference source not found.). Hence, students in the state of engagement/flow are continuously being challenged within their zones of optimal learning (Brown, Ellery, & Campione, 1998; Vygotsky, 1978) and are experiencing two-step episodes alternating between confusion and insight in route to flow/engagement. In contrast to these beneficial flow-confusion-flow cycles, there are the harmful oscillations between boredom and frustration. The harmful oscillation results in disengagement from the task.

Confusion plays a central role in the learning process because it is the gateway to either positive (flow) or negative (frustration) emotions. This is the nexus where individual differences among students undoubtedly have a major role. The positive path is expected from students who
have higher domain knowledge, persistence, mastery orientation, intrinsic motivation, academic risk taking, and willful effort allocation. The negative path is expected from students who have low values on these traits and have the self-concept that they are not talented in the subject matter. Confusion indeed predicts positive learning gains on deep knowledge to the extent that the positive path reigns in the learning environment and student population (Craig, et al., 2004; D’Mello & Graesser, in press; Graesser, D’Mello, et al., 2007). The negative path yields negative correlations with learning gains because hopeless confusion results in frustration, boredom, and disengagement from the task. The vicious cycle of boredom and frustration is very different from the virtuous cycle of confusion and flow.

There is a role for delight and surprise in the emotion transition framework, but these emotions did not occur frequently enough for a reliable analysis. Both of these emotions have short durations so they are often missed in the 20 second polling of our methodology. Some tentative conclusions can be offered from our modest data set. First, surprise can either have a positive or negative valence. Surprise after a sudden insight is positive whereas surprise after unexpected negative feedback is negative. Delight also occurs after a sudden insight (eureka) and when there is positive feedback from AutoTutor, particularly after a difficult challenge is conquered. We would expect more delight emotions for mastery oriented students who experience enlightening conceptual breakthroughs whereas the positive feedback from AutoTutor would engender delight in performance oriented students. These predictions regarding surprise and delight need to be tested more rigorously in future research.
4. Responding to and Eliciting Student Emotions

4.1 Emotion-Sensitive AutoTutor

We recently designed an emotion-sensitive AutoTutor, called *Affective AutoTutor*. Affective AutoTutor automatically detects student emotions based on the multiple channels reported in section 3.1.5 (D’Mello & Graesser, 2010) and responds to the students’ affective-cognitive states by selecting appropriate discourse moves and displaying emotions in facial expressions and speech (D’Mello & Graesser, 2012, in press; D’Mello, Lehman, et al., 2010; D’Mello, Craig, Fike, & Graesser, 2009). The primary student emotions that Affective AutoTutor tries to handle strategically are confusion, frustration, and boredom because these are the emotions that run the risk of leading to disengagement from the task. The tutor continues business as usual when the student is emotionally neutral or in the state of flow/engagement. The emotions of delight and surprise are fleeting, so there is no need to respond to these states in any special way.

The cognitive disequilibrium framework predicts that confusion is a critical juncture in the learning process that is sensitive to individual differences. Some students give up when experiencing confusion because they have a self-concept that they are not good at the subject matter or they try to avoid negative feedback (Dweck, 1999; Meyer & Turner, 2006). Other students treat confusion as a challenge to conquer and expend cognitive effort to restore equilibrium. The first type of student needs encouragement, hints, and prompts to get the student over the hurdle, whereas the second type would best be left to the student’s own devices. An adaptive tutor would treat these students differently. One speculation is that each student has a *zone of optimal confusion* that varies with the student’s background knowledge and interest in the subject matter.
Unlike confusion, responses to frustration would not vary as a function of the student’s level of subject matter knowledge. When the student is frustrated, the tutor would give hints or prompts to advance the student in constructing knowledge and make supportive empathetic comments to enhance motivation.

When the student is bored, the tutor response would once again depend on the knowledge level of the student. The tutor would present more engaging material or challenging problems for the more knowledgeable student. Easier problems are appropriate for the students with low subject matter knowledge so the student can build self-efficacy.

Affective AutoTutor implements most of these strategies in responding to the affective-cognitive states of students. This is accomplished by mechanisms that both detect student emotions and respond in a manner that contributes to student learning. An automated emotion classifier is necessary for Affective AutoTutor to be responsive to learner emotions. As discussed in section 3.1.5, we developed and tested an automated emotion classifier for AutoTutor based on the dialogue history, facial action units, and position of student’s body during tutoring (D’Mello & Graesser, 2010). The features from the various modalities can be detected in real time automatically on computers, so we have integrated these sensing technologies with Affective AutoTutor.

An emotion generator was also needed for Affective AutoTutor because the system was expected to respond with suitable emotions. Therefore, the agent needed to speak with intonation that was properly integrated with facial expressions that displayed emotions. There was an enthusiastic nod with positive feedback language to be used after the student has a correct contribution. There was a shaking of the head with a skeptical look when the student contribution was low quality. There was an empathetic expression conveyed in words, facial
expressions, and motion when supportive encouragement would be needed. A small set of emotion displays like these examples went a long way in conveying the tutor’s emotions.

It is too early to make any firm conclusions about the impact of Affective AutoTutor on learning, but we have conducted some studies. We compared the original AutoTutor without emotion tracking and emotional displays to an Affective AutoTutor version that is emotionally supportive. The supportive Affective AutoTutor had polite and encouraging positive feedback (“You’re doing extremely well”) or negative feedback (“This is difficult for most students”). When the student expressed low quality contributions, the tutor attributed the problem to the difficulty of the materials and it being challenging for most students rather than blaming the student. There was another shake-up version of Affective AutoTutor. This version tried to shake up the emotions of the student by being playfully rude and telling the student what emotion the student is having (“I see that you are frustrated”). Instead of giving earnest feedback, the shake-up AutoTutor gave positive feedback that is sarcastic (e.g., “Aren’t you the little genius”) and negative feedback that is derogatory (e.g., “I thought you were bright, but I sure pegged you wrong”). The simple substitution of this feedback dramatically changes AutoTutor’s personality. The shake-up tutor is very engaging for some students whereas other students would prefer to interact with the polite supportive tutor.

The data we have collected revealed that the impact on learning from the different tutors appears to depend on the phase of tutoring and the student’s level of mastery. An emotion-sensitive AutoTutor had either no impact or a negative impact on learning during early phases of the tutoring session. During a later phase of tutoring, the polite supportive AutoTutor improved learning, but only for the low knowledge students. Although more studies need to be conducted, it is tempting to speculate that emotional displays by AutoTutor may not be beneficial during the early phases of
an interaction when the student and agent are “bonding” and that a supportive polite tutor is appropriate at later phases for students who have low knowledge and encounter difficulties. Perhaps the playful shake-up tutor is motivating when boredom starts emerging for the more confident, high-knowledge learners. These conclusions are quite tentative, however, because there needs to be more research in diverse student populations and learning environments.

4.2 Planting Cognitive Disequilibrium

The claim has been made throughout the chapter that cognitive disequilibrium gives rise to confusion and enhances learning of difficult material. We documented that learning gains were positively correlated with confusion as long as the learner was not hopelessly confused (Craig et al., 2004; D’Mello & Graesser, 2012; Graesser, D’Mello, Chipman, King, & McDaniel, 2007). The question arises, however, whether there is a causal relationship between (a) cognitive disequilibrium and (b) confusion and/or learning. We have conducted some studies that manipulated cognitive equilibrium experimentally and measured the consequences on confusion and learning (Lehman et al., 2011).

Lehman et al. (2011) used the case study modules and trialogs in the ARIES system (Millis et al., in press) to systematically manipulate cognitive disequilibrium. We did this by manipulating whether or not the tutor agent and the student agent contradicted each other during the trialog and expressed points that are incorrect. Each case study had a description of a research study that was to be critiqued during the trialogs. In the True-True control condition, the tutor agent expressed a correct assertion and the student agent agreed with the tutor. In the True-False condition, the tutor expressed a correct assertion but the student agent disagreed by expressing an incorrect assertion. In the False-True condition it was the student agent who
provided the correct assertion and the tutor agent who disagreed. In the *False-False* condition, the tutor agent provided an incorrect assertion and the student agent agreed.

The human student was asked to intervene after each point of possible contradiction; the agents turned to the human and asked “So what would your decision be, <name of student>”? If the human student experiences uncertainty and is confused, this should be reflected in the incorrectness/uncertainty of the human’s answer. Uncertainty is a likely opportunity to scaffold deep comprehension by forcing learners to stop and think. The data indeed confirmed that the contradictions and false information had an impact on the humans’ answers to these questions. The probability of their giving a correct answer to a binary decision (chance = .50) question immediately following a contradiction was .76, .60, .45, and .35 in the True-True, True-False, False-True, and False-False conditions, respectively. Uncertainty/incorrectness is low when both agents are correct and there is no contradiction (*true-true*), but increases when one agent of the agents is incorrect. Uncertainty is greater when the tutor is incorrect (*false-true*) compared to when the tutor is correct (*true-false*), presumably because the former situation is incompatible with conventional norms. Uncertainty is greatest when both agents are incorrect, even without a contradiction (*false-false*). This can be explained by either conformity of the human with the two agents or the detection of a clash between the human student’s knowledge and the agents’ responses. We suspect that the former is most likely.

Confusion would presumably be best operationally defined as occurring if both (a) the student identified the experience as confusion and (b) the student manifests uncertainty/incorrectness in their decisions when asked by the agents. Lehman et al.’s analysis of retrospective emotion judgments by the human student unfortunately had a very low rate of reported confusion (5%) among the various conditions so the measure was not sensitive. The
automated measures of confusion detection would be more sensitive but these were not reported in Lehman et al. (2011).

So far there is some evidence that manipulated cognitive disequilibrium causes an increase in uncertainty and allegedly confusion. Is there any evidence that disequilibrium and/or confusion causes more learning at deeper levels of mastery? A delayed test on scientific reasoning sheds some light on this question. The results indicated that contradictions in the false-true condition produced higher performance on multiple choice questions that tapped deeper levels of comprehension than performance in the true-true condition. These data suggests that the most uncertainty occurs when the tutor makes false claims that student agents disagree with. This contradiction stimulates thought and reasoning at deeper levels. Comprehension scores on a delayed posttest are improved by this experience. These data suggest there may be a causal relationship between cognitive disequilibrium and deep learning, with confusion playing either a mediating or moderating role in the process. More research is obviously needed to dissect the timing and causal status of disequilibrium, confusion, and deep learning.

5. Conclusions

This chapter has reviewed our program of research that investigates moment-to-moment emotions that occur during complex learning. We have discovered a number of novel findings in our exploration of learning during conversational interactions with AutoTutor and other learning environments. The learning-centered emotions are confusion, frustration, boredom, and flow-engagement, with occasional moments of delight and surprise. Anxiety also occurs when students face high-stakes tests and curiosity occurs when intrinsic motivation is high. With the exception of surprise, these are not the emotions of Ekman’s big six. These emotions can be identified by the student, peers, teachers, and trained experts, but the different judges show modest agreement, except
for trained judges on the more salient emotions that are manifested on the face. There are automated methods of classifying emotions from the channels of tutorial dialogue history, facial expressions, and body posture. For some emotions (confusion, delight, surprise) the face is most diagnostic, whereas body posture is most diagnostic for others (boredom, flow/engagement) and dialogue history is needed to detect frustration. A combination of these channels yields the best classifier, which yields emotion detection performance that exceeds the novice judges but is a bit shy of the experts. The duration of emotions is longer for boredom, flow, and confusion than delight and surprise, with frustration in between. The occurrence and transitions between emotions is explained reasonably well by a cognitive disequilibrium theoretical framework. The student experiences impasses that trigger cognitive disequilibrium and confusion. Confusion might be resolved and equilibrium is restored; alternatively, unresolved confusion and persistent failure leads to frustration and boredom. There appears to be a causal relationship between cognitive disequilibrium and confusion, which in turn leads to thoughtful reasoning and deeper learning. Attempts to manipulate cognitive disequilibrium by contradictions between agents (tutor and student) were successful and sometimes causally yielded higher scores on multiple choice tests. We created an emotion sensitive AutoTutor that both detects the emotions of the learner and responds with a tutor agent that displays emotions and tries to promote deep learning. The emotion sensitive AutoTutor does produce increases in learning gains, but it depends on the subject matter knowledge of the student and the phase in the tutoring session.

We believe our research program has uncovered new ground in understanding the dynamic relationships between cognition and emotion. Very few researchers have performed fine-grained technologically-assisted investigations of the emotions that occur during complex learning. This is
somewhat of an amazing oversight in a country that worries so much about how we can improve motivation for students to learning difficult STEM topics.

The research discussed in this chapter provides an initial sketch of emotions during complex learning. Quite clearly, more research is needed on so many levels. It is important to document the emotions, durations of emotions, and transitions between emotions in a diverse array of learning environments. The cognitive disequilibrium framework is a good start but there needs to be a systematic investigation of how components in the framework are influenced by individual differences among students with respect to subject matter knowledge, general reasoning skills, academic risk taking, intrinsic motivation, persistence, emotional intelligence, self-concept, the list goes on. These individual differences are undoubtedly mediating or moderating variables in the system, but this needs to be documented. Available evidence suggests that confusion is a pivotal emotion that sometimes leads to deeper learning, but there is uncertainty on how to scale the zone of optimal confusion for individual students. There are also questions on how to design AutoTutor and other learning environments in a manner that responds to student emotions appropriately in ways that promote deep learning.

The links between emotions and deep learning emerge in the design of serious games. Emotions are of course central to the design of educational games (Conati, 2002; McNamara, Jackson, & Graesser, in press; McQuiggan, Mott, & Lester, 2008; Millis et al., in press; Moreno & Mayer, in press; Shaffer, 2006). Educational games ideally are capable of turning work into play by minimizing boredom, optimizing engagement/flow, presenting challenges that reside within the optimal zone of confusion, preventing persistent frustration, and engineering delight and pleasant surprises. The design of the ideal serious game would be a perfect, perhaps lucrative application of the science we have discussed in this chapter.
Acknowledgements

The research on was supported by the National Science Foundation (ITR 0325428, REESE 0633918, ALT-0834847, DRK12-0918409, DRL-1108845) and the Institute of Education Sciences (R305A080594, R305G020018). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding sources.
References


Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work* (pp. 16–23). Amsterdam: IOS Press.


Computers for Mathematical Learning 9, 169-211.


Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 574-586). New York: Guilford Press.


Table 1: Kappa inter-judge reliability scores for affective states at all points, mandatory points, and voluntary points.

<table>
<thead>
<tr>
<th>Pair of Judges</th>
<th>All</th>
<th>Mandatory</th>
<th>Voluntary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self/Peer</td>
<td>0.08</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>Self/Expert1</td>
<td>0.14</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Self/Expert2</td>
<td>0.16</td>
<td>0.13</td>
<td>0.24</td>
</tr>
<tr>
<td>Peer/Expert1</td>
<td>0.14</td>
<td>0.11</td>
<td>0.36</td>
</tr>
<tr>
<td>Peer/Expert2</td>
<td>0.18</td>
<td>0.15</td>
<td>0.37</td>
</tr>
<tr>
<td>Expert1/Expert2</td>
<td>0.36</td>
<td>0.31</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Figure Caption Page

Figure 1. Cognitive Disequilibrium Framework.

Figure 2. (A) AutoTutor Interface and (B) Sample Dialogue from a Tutorial Session.

Figure 3. Automated Sensing of Emotions.

Figure 4. Examples of Affective States.

Figure 5. Coded Emotions as a Function of Time in the AutoTutor Session.
**A) AutoTutor Interface**

When you turn on the computer, how is the operating system first activated and loaded into RAM?  

**Student Response:**

**B) Sample Tutorial Dialogue**

**T:** When you turn on the computer, how is the operating system first activated and loaded into RAM?  
**[Main question]**

**S:** When the computer is turned on a file is automatically booted up.

**T:** Anything else?  
**[Pump for info]**

**S:** This file contains the operating system.

**T:** What about storage?  
**[Hint]**

**S:** The operating system is then read into RAM and activated.

**T:** Right. The CPU executes instructions in ROM, which helps the operating system to start.  
**[After positive feedback, there is an assertion]**
A. Confusion  
B. Delight  
C. Boredom  
D. Frustration