

Running head: CONFUSION AND COMPLEX LEARNING

Confusion and Complex Learning during Interactions with Computer Learning Environments

Blair Lehman, Sidney D'Mello, & Art Graesser

University of Memphis

Contact person:

Blair Lehman,

202 Psychology Building,

The University of Memphis,

Memphis, TN 38152.

Phone: 847-942-3789.

Email: balehman@memphis.edu

Abstract

Folk wisdom holds that being confused is detrimental to learning. However, research on emotions and learning suggest a somewhat more complex relationship between confusion and learning outcomes. In fact, it has been proposed that impasses that trigger states of cognitive disequilibrium and confusion can create opportunities for deep learning of conceptually difficult content. This paper discusses four computer learning environments that either naturally or artificially induce confusion in learners in order to create learning opportunities. First, an Intelligent Tutoring System called AutoTutor that engenders confusion through challenging problems and vague hints is described. The remaining three environments were specifically designed to induce confusion through a number of different interventions. These interventions include device breakdowns, contradictory information, and false feedback. The success and limitations of confusion induction and the impact of confusion resolution on learning are discussed. Potential methods to help learners productively manage their confusion instead of being hopelessly confused are also discussed.

Keywords: confusion, cognitive disequilibrium, complex learning, Intelligent Tutoring Systems, confusion induction

Confusion and Complex Learning during Interactions with Computer Learning

Environments

1. Introduction

Complex learning is an emotionally charged experience. Complex learning occurs when learners work towards comprehending difficult material, solve a difficult problem, or make a difficult decision (Chi, 1992; Chi & Ohlsson, 2005; Graesser, Ozuru, & Sullins, 2010). Throughout this process there is a natural ebb and flow between positive and negative emotions, coinciding with the struggles and successes that learners experience during effortful problem solving, reasoning, and comprehension.

In traditional learning environments, such as classrooms or human-human tutoring sessions, instructors are responsible for adapting to both the cognitive and affective states of learners as they face challenges and work to maintain motivation. This education model is gradually changing, however, as non-traditional learning environments are becoming increasingly prevalent in the 21st century. For example, Google and the Internet are replacing the library as “go to” sources for information (Anderson, 2004). Intelligent Tutoring Systems (ITSs) and other forms of computer-based training are learning resources that have become more prevalent in the 21st century (Graesser, VanLehn, Rose, Jordan, & Harter, 2001; Koedinger & Corbett, 2006; Psozka, Massey, & Mutter, 1988; VanLehn et al., 2007; Woolf, 2009). This trend of moving away from traditional learning environments has decreased the prevalence of human-human interactions during learning, which dominated the educational landscape for millennia. Despite this reduction in human-human interactions, students still experience a variety of positive and negative emotions during learning (Calvo & D’Mello, 2011; Meyer & Turner, 2006;

Pekrun, 2010; Schutz & Pekrun, 2007). Hence, much like a gifted human mentor (Goleman, 1995; Lepper & Woolverton, 2002), it is important that these new learning technologies intelligently manage learner emotions.

Fifty years ago, researchers viewed emotion and cognition as either separate or competing processes with respect to successful decision making, problem solving, and learning (Damasio, 1994). This dichotomy between the rational and the emotional mind has been challenged by a preponderance of research that has shown that emotions and cognition are inextricably linked (Barrett, Mesquita, Ochsner, & Gross, 2007; Bower, 1981; Csikszentmihalyi, 1990; Dagleish & Power, 1999; Lazarus, 1999; Mandler, 1984; Ortony, Clore, & Collins, 1988; Scherer, Schorr, & Johnstone, 2001; Stein & Levine, 1991). Research now shows that emotions operate continually throughout cognitive processes such as memory, reasoning, problem solving, and deliberation (Mandler, 1984; Stein, Hernandez, & Trabasso, 2008). Importantly, there is considerable overlap in the neural circuitry that supports cognitive and emotional processes (Dagleish, Dunn, & Mobbs, 2009; Immordino-Yang & Damasio, 2007; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, in press), further suggesting that cognition and emotion are two sides of the same coin (Lazarus, 2000).

Although the 21st century has been ripe with theory exploring links between emotions and cognition, with the exception of test anxiety (Pekrun, Goetz, Daniels, Stupinsky, & Raymond, 2007; Zeidner, 2007), investigations into the experience and impact of emotions on learning are much more sparse and scattered. The learning of technical and conceptual subject matter is difficult and learners inevitably experience obstacles, failures, and the resultant negative emotions such as anxiety, confusion, and frustration. Positive emotions such as delight, pride, and satisfaction, are also experienced when challenges are conquered and important goals

are achieved (Ahmed, van der Werf, Minnaert, & Kuyper, 2010; Dettmers et al., 2011; Pekrun, 2011; Smith & Kirby, 2009).

The emotions that learners experience during a broad range of educational activities are sometimes called *academic emotions* (Pekrun, 2010). Pekrun and colleagues have classified the academic emotions into four main categories: achievement, topic, social, and epistemic. These four categories cover the wide range of emotions that learners experience during learning activities over the course of an entire semester or year of school. Learners experience emotions with regard to (a) outcomes (*achievement*, e.g., contentment, anxiety, and frustration), (b) preferences for certain topics over others (*topic*, e.g., empathy for the protagonist in a novel), (c) interactions with peers and teachers (*social*, e.g., pride, shame, and jealousy), and (d) processing new information that is encountered (*epistemic*, e.g., surprise and confusion).

In contrast to this broad set of academic emotions, research has also focused on in depth analyses of shorter learning sessions, lasting from 30 minutes to 1.5 hours (Arroyo, et al., 2009; Burleson & Picard, 2007; Chaffar, Derbali, & Frasson, 2009; Conati & Maclaren, 2009; D’Mello, Craig, Fike, & Graesser, 2009; Forbes-Riley & Litman, 2009; Robison, McQuiggan, & Lester, 2009). Fine-grained analyses on these shorter learning sessions indicated a different set of emotions, consisting of the academic emotions along with some other emotions. These “learning-centered” emotions include anxiety, boredom, confusion, curiosity, engagement/flow, frustration, happiness, delight, and surprise (Calvo & D’Mello, 2011; Rodrigo & Baker, 2011a). This paper focuses on these shorter learning periods and the set of learning-centered emotions.

In addition to variations in the duration of learning (e.g., semester vs. 30 minutes to 1.5 hours), the learning environment can also be varied. Research investigating the learning-centered emotions has looked at learning environments such as expert human tutoring sessions (Lehman,

Mathews, D'Mello, & Person, 2008), individual problem solving (D'Mello, Lehman, & Person, 2010), and ITSs (Baker, D'Mello, Rodrigo, & Graesser, 2010; Craig, Graesser, Sullins, & Gholson, 2004; D'Mello, Craig, Sullins, & Graesser, 2006; D'Mello & Graesser, 2011; Graesser, Chipman, King, McDaniel, & D'Mello, 2007; McQuiggan, Robison, & Lester, 2010).

Educational games have also become an effective medium for learning and are likely to bring about a variety of learner emotions. But, how do learners' emotional experiences differ when learning from an educational game compared to an ITS? Rodrigo & Baker (2011b) investigated this question by comparing learner emotions when learning math from an ITS, Aplusix, and an educational game, Math Blaster 9-12. Experiences of confusion, frustration, and surprise did not differ between the ITS and educational game. However, learners were found to experience lower levels of boredom and delight, and higher levels of engagement/flow when interacting with the ITS compared to the educational game.

Learner emotions have also been studied in Crystal Island, an educational game that covers microbiology and genetics (McQuiggan et al., 2010). Learner self-reports showed that learning-centered emotions occurred frequently, with engagement/flow, confusion, and delight occurring most frequently. In addition, learners also reported experiencing excitement during the interaction. These findings suggest that while learning activities in general induce emotions, the learning environment can impact which emotions are experienced.

As well as identifying when an emotion occurs, it is also possible to analyze the context surrounding the emotional episode. One approach involves building computational models of learners' emotional experiences to inform emotionally aware agents about the current state of the learner. Conati & McLaren (2009) have constructed dynamic Bayesian networks based on the OCC model of emotions (Ortony et al., 1988), which is an appraisal theory of emotion. Their

model allows the emotionally aware agent to not only predict the emotion that the learner is expected to be experiencing, but also affords a form of causal reasoning to understand why the learner is experiencing that emotion.

Some research has also focused on correlating the emotions that arise during a learning session with learning outcomes. Boredom was found to be negatively correlated with learning outcomes, while engagement/flow was positively related to learning (Baker et al., 2010; Craig et al., 2004; D'Mello et al., 2006; D'Mello & Graesser, 2011; Graesser, Chipman et al., 2007). While the correlations involving boredom and engagement/flow are in the expected directions, perhaps the most intriguing finding is the positive correlation between confusion and learning. More specifically, confusion was significantly correlated with learning gains in three studies involving tutoring sessions with AutoTutor, an ITS with conversational dialogues (Craig et al., 2004; D'Mello & Graesser, 2011; Graesser, Chipman et al., 2007).

Confusion has also been found to be both prevalent and important to learning with ITSs (Baker et al., 2010; Craig et al., 2004; D'Mello & Graesser, 2011; Graesser, Chipman et al., 2007; Rodrigo & Baker, 2011a), and even in the absence of technology (Lehman et al., 2008; D'Mello, Lehman, & Person, 2010; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). In one study of 50 hours of expert human tutoring, Lehman et al. (2008) found that confusion was the most frequently occurring emotion when compared to both "basic" emotions (e.g., fear, anger, sadness, disgust) (Ekman, 1973) and other learning-centered emotions. Confusion was also found to frequently occur when learners both asked and answered questions during the tutoring session (Lehman, D'Mello, & Person, 2010). Impasses, which are states that potentially trigger confusion, were found to play an important role in VanLehn et al.'s (2003) analysis of over 100 hours of human-human tutoring sessions. Their results indicated that deep learning of

physics concepts rarely occurred when learners did not reach an impasse, irrespective of the quality of explanations provided by the tutor.

In summary, the apparent benefits of impasses and the resultant confusion beg the question: How can learning environments take advantage of confusion to increase student learning? The present paper addresses several aspects of this question by synthesizing one line of research that focuses on confusion. It should be emphasized that the purpose of this paper is not to conduct a full review of the literature on confusion, but rather to synthesize findings from four studies that highlight the importance of confusion during learning with technology.

The remainder of this paper is divided into four sections. First, a review of some of the theories on the emergence and resolution of confusion are presented (Section 2). Next, studies with four computer learning environments are discussed that attempt to *productively* confuse learners (Section 3). Third, possible interventions to help learners productively manage their confusion are discussed (Section 4). The paper concludes by discussing future research directions (Section 5).

2. Theories on the Emergence, Resolution, and Impact of Confusion

The positive relationship between confusion and learning is consistent with theories that emphasize the importance of goal appraisal as an antecedent to emotion (Stein & Levine, 1991), with theories that highlight the merits of impasses during learning (Brown & VanLehn, 1980; VanLehn et al., 2003), and with theories that claim that cognitive disequilibrium (described below) is one precursor to deep learning (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Graesser & Olde, 2003). According to these theories, confusion is triggered when learners are confronted with information that is inconsistent with existing knowledge and learners are unsure

about how to proceed. These events that trigger impasses place learners in a state of cognitive disequilibrium, which is ostensibly associated with heightened physiological arousal and more intense thought as learners attempt to resolve impasses. Importantly, these theories posit that confusion itself does not cause learning gains, but the cognitive activities that accompany confusion, cognitive disequilibrium, and impasse resolution are linked to learning. These theories are briefly described below.

When applied to learning contexts, goal-appraisal theory focuses on the importance of learning progress in relation to learning goals when predicting the learner's emotional state (Scherer et al., 2001; Stein & Levine, 1991). This theory emphasizes appraisals of the relevance of events to current goals and plan availability. During a complex learning task, learners generate goals and sub-goals that they are motivated to achieve. For example, a learner may have the superordinate goal of getting an A on an upcoming biology test, and might generate sub-goals to understand cells, reproduction, ecosystems, etc. Learners will inevitably encounter events that either facilitate or inhibit both sub- and superordinate goals when working towards achieving their superordinate goal (e.g., getting an A).

Events that facilitate goal achievement are expected to produce positively-valenced emotions (i.e., happiness) and inhibitory events are expected to produce negatively-valenced emotions (i.e., frustration). The intensity of the emotional experience, though, is determined by the degree to which an event impacts goal achievement. An event such as answering a question correctly during a study session may bring about feelings of satisfaction or mild happiness, whereas receiving an A on an exam would be more likely to induce a more pronounced positive emotion such as joy.

Events that facilitate goal achievement vary substantially in intensity, but goal-blocking events vary in more intricate ways because a plan is needed when goals are blocked. When a plan is available, the goal-blocking event is likely to produce a state of mild irritation because it can be resolved by instantiating the plan. However, when a plan is not available, the obstacle is likely to evoke stronger negative emotions. Frustration is expected to occur in a situation where a learner hits an obstacle, perceives the obstacle to be highly relevant to the current goal, and perceives no available plan to continue toward the goal.

In addition to the goal-related event influencing learners' emotional experiences, the current emotional state can also influence goal achievement. In the case of a goal-blocking event, the current emotional state can influence the ability of learners to formulate an effective plan to move past the goal-blocking event. For example, as frustration becomes more intense, learners will become less capable of discovering a new strategy to achieve their goal, in part because of the detrimental effects of negative emotions (e.g., frustration) on cognitive processes such as working memory, attention, and creativity (Clore & Huntsinger, 2007; Forgas, 1995; Isen, 2008).

Of particular relevance to the current paper are obstacles to goals that trigger impasses and confusion. An impasse may occur when there is a clash between prior knowledge and incoming information (such as a contradiction, anomaly, system breakdown, or error) and when there is uncertainty about how to proceed, thereby triggering a state of cognitive disequilibrium. Cognitive disequilibrium is a state of uncertainty that occurs when an individual is confronted with obstacles to goals, interruptions of organized action sequences, impasses, contradictions, anomalous events, dissonance, incongruities, unexpected feedback, uncertainty, deviations from norms, and novelty (Bjork & Linn, 2006; D'Mello & Graesser, in press; Festinger, 1957;

Graesser et al., 2005; Piaget, 1952). Confusion is hypothesized to be the affective component of cognitive disequilibrium.

An important hypothesis is that confusion can mediate learning activities or even cause the learner to begin effortful problem solving and careful deliberation in order to restore equilibrium. These effortful cognitive activities can result in deeper processing on the part of the learner, and thereby the impasses and confusion can lead to increased opportunities for learning to occur (Brown & VanLehn, 1980; Carroll & Kay, 1988; Graesser & Olde, 2003; Siegler & Jenkins, 1989; VanLehn et al., 2003). However, these learning opportunities do not guarantee that learners will acquire a better understanding of the topic. Impasse-driven theories of learning propose that it is the successful resolution of an impasse and not the impasse itself that is associated with learning (D'Mello & Graesser, in press; VanLehn et al., 2003).

3. Interventions to Induce Confusion

Recent research has shown that confusion is both a prevalent emotion during learning and is positively correlated with learning, particularly at deeper levels of comprehension (Baker et al., 2010; Craig et al., 2004; D'Mello & Graesser, 2011; D'Mello, Lehman, Sullins et al., 2010; Graesser, Chipman et al., 2007; Lehman et al., 2008). In addition, cognitive disequilibrium theory and impasse-driven theories of learning suggest that experiences of confusion provide opportunities for deeper learning. In this vein, three computer learning environments that were specifically designed to trigger confusion during learning have been developed. Through learner interactions with these three environments, it was possible to investigate the conditions under which confusion could be successfully induced as well as how the experiences of confusion impacted learning outcomes. First, an ITS, AutoTutor, is discussed. This learning environment

was not designed to trigger confusion, but natural experiences of confusion were found to occur during learning sessions with AutoTutor.

3.1 Learning of Computer Literacy with AutoTutor

AutoTutor. AutoTutor is an ITS with mixed-initiative dialogue that teaches learners difficult topics such as Newtonian physics, computer literacy, and critical thinking. AutoTutor's dialogues are organized around difficult questions and problems (called main questions) that require reasoning and explanations in the answers. When presented with these questions, learners typically respond with answers that are only one word to two sentences in length. In order to guide learners in their construction of an improved answer, AutoTutor actively monitors learners' knowledge states and engages them in a turn-based dialogue. AutoTutor adaptively manages the tutorial dialogue by providing feedback on the learner's answers (e.g., "good job", "not quite"), "pumping" the learner for more information (e.g., "What else?"), giving hints (e.g., "What about X?"), giving prompts (e.g., "X is a type of what?"), correcting misconceptions, answering questions, and summarizing topics.

Learning gains produced by AutoTutor have ranged from 0.4-1.5 sigma (a mean of 0.8), depending on the learning measure, the comparison condition, the subject matter, and the version of AutoTutor (Kopp, Britt, Millis, & Graesser, in press; Graesser et al., 2004; VanLehn et al., 2007). A 1 sigma effect size is approximately a one letter grade increase in learning.

AutoTutor's effectiveness stems from its dialogue management system (see Table 1 for dialogue excerpt), which is modeled after the dialogue patterns of human tutors (Graesser, Person, & Magliano, 1995). Learners engage in an interactive dialogue with an animated conversational agent utilizing speech, facial expressions, and some rudimentary gestures (see Figure 1). Each tutoring interaction begins with a difficult question that requires three to four

sentences to answer correctly (see turn 1 in Table 1 for a sample physics question). Learners generate an initial response to this difficult question (see turn 2). After assessing learners' initial responses to the main question, AutoTutor begins a collaborative interaction with learners to target gaps in knowledge (expectations that were not covered in learners' initial responses) and to identify and correct misconceptions (see turns 3-18). This is accomplished by comparing learners' responses to a set of expectations and misconceptions from a curriculum script. This comparison is accomplished through matching operations based on symbolic interpretation algorithms (Rus & Graesser, 2007) and statistical semantic matching algorithms (Graesser, Penumatsa, Ventura, Cai, & Hu, 2007).

Insert Figure 1 About Here

Insert Table 1 About Here

Confusion during learning with AutoTutor. Learner emotions during interactions with AutoTutor have been investigated in a number of studies using a variety of methods to monitor emotions (Craig et al., 2004; D'Mello et al., 2006; Graesser, Witherspoon, McDaniel, D'Mello, Chipman, & Gholson, 2006; Graesser, Chipman et al., 2007; Pour, Hussein, Al Zoubi, D'Mello, & Calvo, 2010). In these studies, learners interacted with AutoTutor for 30 minutes to 1.5 hours on various topics in computer literacy (hardware, operating system, Internet). Confusion was measured with a number of methods including online observations, emote-aloud protocols, cued-recall, and coding of video data by peers and trained judges.

Confusion was measured using three different types of judgments: self, peer, and trained judges. However, only the Graesser et al. (2006) study utilized all three types of judges. Self-judgments were completed by the learners, peer-judgments were completed by a second learner that had interacted with AutoTutor, and the trained judges were researchers that had been trained

on the facial action coding system (FACS; Ekman & Friesen, 1978). With the exception of one study (Craig et al., 2004), confusion judgments were made retrospectively, after the interaction with AutoTutor had been completed. In Craig et al. (2004) learner emotions were judged by four trained human judges. Judges conducted online observations once every 5 minutes and noted which emotion the learner was experiencing. The retrospective affect judgment protocol (Graesser et al., 2006) involved judges (self, peer, trained) watching videos of learners' faces and computer screens during the interaction with AutoTutor. Judges were given a list of six affective states (confusion, frustration, boredom, engagement/flow, delight, surprise) and neutral, along with definitions, to select from when making judgments. Judgments could be made spontaneously and at pre-specified points in the learning interaction. Pre-specified judgments were either specific time intervals (e.g., every 20 seconds; Graesser et al., 2006; Pour et al., 2010) or specific points in the learning interaction (e.g., after feedback, before a learner contribution; Graesser, Chipman et al., 2007).

When averaged across five studies, the following distribution of emotions has been found: **confusion (17%)**, frustration (13%), boredom (18%), engagement/flow (24%), delight (6%), surprise (3%), and neutral (19%). Indeed, confusion is quite prevalent during interactions with AutoTutor.

Although confusion was not experimentally induced during interactions with AutoTutor, there are several aspects of AutoTutor's pedagogical strategies that engender confusion. In particular, confusion is expected to occur due to the difficult problems or questions that AutoTutor presents to the learners. The questions required answers that involved inferences and deep reasoning, such as *why*, *how*, *what-if*, *what if not*, and *how is X similar to Y*? Examples of these questions are, "How can John's computer have a virus but still boot to the point where the

operating system starts?” (hardware question) and “How will you design a network that will continue to function, even if some connections are destroyed? (Internet).”

The vagueness of AutoTutor’s hints and pumps is also expected to trigger confusion and deep thinking. These vague hints and pumps are expected to cause learners to confront potential gaps in their knowledge and to begin effortful deliberation in order to construct a better mental model of the material.

There is some evidence to support these hypotheses. D’Mello and colleagues mined AutoTutor’s log files and examined the tutorial dialogue (i.e., the context) over 15 second intervals that culminated in episodes of confusion (D’Mello et al., 2006; D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008). An event triggering an emotional episode could either be tutor generated (i.e., boredom because the tutor is providing a long-winded explanation), learner generated (i.e., boredom because the learner has no interest in computer literacy), or a session related event (i.e., boredom because the tutorial session is dragging on). It was discovered that confusion occurred earlier in the session, within the first few attempts to answer a question, with slower and less verbose responses, with poor answers, with frozen expressions (instead of domain related contributions), when the tutor is less direct (hints and pumps), and when the tutor provides negative feedback. As a point of comparison, boredom occurred later in the session, after multiple attempts to answer a question, and when AutoTutor gives more direct dialogue moves (i.e., assertions or summaries that are more direct than pumps or hints).

In summary, over many studies, AutoTutor has been found to both promote confusion and learning in a tutoring context. Confusion generated during AutoTutor’s learning sessions has been positively correlated with learning (Craig et al., 2004; D’Mello et al., 2006; Graesser,

Chipman et al., 2007). This suggests that implementations of particular pedagogical strategies (e.g., difficult problems, hints, pumps) can induce confusion, which if productively managed, can result in improved learning.

3.2 Breakdown Scenarios during Device Comprehension

This set of experiments investigated the possibility of inducing confusion while learners completed a text-diagram comprehension task (D'Mello & Graesser, in review). Specifically, learners were provided with explanatory texts and diagrams of every day devices (e.g., toaster, cylinder lock, etc.), from Macauley's (1988) *The Way Things Work*, and asked to construct a mental model of how the devices functioned.

Participants and design. The participants in the two experiments were undergraduate students from a large mid-South university in the United States. There were 52 participants in the first experiment and 88 participants in the second experiment. In both experiments a within-subjects design was used. Participants were presented with four devices, two devices were studied in the experimental condition and two were studied in the control condition.

Manipulation and interaction. In the two experiments, learners were first presented with an illustrated text of a device and instructed to comprehend how the device worked. The illustrated text (Macauley, 1988) contained an image of an everyday device along with an explanation of how the device functioned. After this initial presentation and study phase, learners were then presented with the same illustrated text along with a description of a device breakdown (breakdown trials). The breakdown described a situation in which the device had stopped functioning. The cylinder lock, for example, had the following breakdown: "A person puts the key into the lock and turns the lock but the bolt doesn't move. Try to understand what is wrong with the cylinder lock".

The control trials involved either re-reading the original illustrated text (Experiment 1) or focusing on a key component of the device while re-reading (Experiment 2). Learners studied four devices in each study across two breakdown trials and two control trials.

Measuring confusion and results. Confusion was measured through two self-report methods. For both confusion measurements, larger scores indicate greater levels of confusion. The first measurement involved coarse-grained online reports of confusion through a questionnaire. In Experiment 1, paired-samples *t*-tests revealed that there were no significant differences in self-reported confusion between the breakdown ($M = 3.10$, $SD = 1.51$) and control trials ($M = 2.83$, $SD = 1.39$), although the trend was in the expected direction. However, in Experiment 2 learners reported higher levels of confusion for the breakdown trials ($M = 2.77$, $SD = 1.40$) compared to the control trials ($M = 2.37$, $SD = 1.24$).

In addition to confusion, learners were also required to report their level of engagement and frustration. In both experiments, there were not significant differences between breakdown and control trials for engagement and frustration when paired-samples *t*-tests were conducted. The one exception was that in Experiment 1, learners reported higher levels of engagement for the breakdown trials ($M = 3.46$, $SD = 1.57$) compared to the control trials ($M = 3.10$, $SD = 1.52$).

The second, finer grained measurement of confusion involved a retrospective confusion judgment protocol (Graesser et al., 2006). These judgments occurred offline, that is, after learners studied all four devices and breakdowns. During the retrospective judgment protocol learners viewed a video of their faces along with the context of the face video (i.e., the specific illustrated text they were viewing at that time). Learners then made continuous confusion judgments on a scale of 0 to 10. These fine-grained confusion ratings also allowed for a temporal

analysis of confusion dynamics over the course of studying the devices. Analyses of the dynamics of confusion revealed two main patterns: partially-resolved and unresolved confusion.

Measuring learning and results. Learners completed a device comprehension test after viewing all four devices. This test assessed learners' comprehension of the functioning of the devices as a whole. Learning outcomes were investigated with respect to confusion resolution. Impasse-driven learning theories would predict that learners that had partially-resolved confusion would perform better than the unresolved group on tests of device knowledge. The results from independent-samples *t*-tests supported this conclusion. Learners who partially-resolved their confusion performed better on the comprehension test (Experiment 1: $M = .48$, $SD = .15$; Experiment 2: $M = .51$, $SD = .17$) than those who were unsuccessful at resolving their confusion (Experiment 1: $M = .38$, $SD = .16$; Experiment 2: $M = .43$, $SD = .16$). The mean effect size was .58 across both experiments, which is consistent with a medium sized effect (Cohen, 1992).

Overall, these two experiments indicated that confusion can be induced through device breakdowns. Moreover, there were differences in device comprehension scores between learners who partially-resolved their confusion compared to their hopelessly confused counterparts. The two experiments did not include an intervention to help learners productively manage their confusion. Instead, confusion resolution was left to the learners' own abilities. Investigations of individual differences revealed that learners with higher scholastic aptitude (measured via self-reported ACT and SAT scores) were more likely to partially-resolve their confusion and demonstrate better device comprehension than the less gifted learners.

3.3 Contradictory Information during the Learning of Critical Thinking Skills

Cognitive disequilibrium theory posits that contradictions are one possible trigger of confusion. This prediction was tested in an experiment that attempted to experimentally induce

confusion via a contradictory information manipulation (Lehman et al., 2011). Contradictory information was presented by two animated pedagogical agents while discussing critical thinking and scientific reasoning topics with human learners.

Participants and design. The participants in this experiment were 32 undergraduate students recruited from a large mid-South university in the United States. There were four conditions in this experiment (described further below): *true-true*, *true-false*, *false-true*, and *false-false*. A within-subjects design was used and all learners completed two learning sessions on different scientific reasoning topics in each condition, for a total of eight learning sessions. Order of conditions and assignment of topic to condition was counterbalanced across learners with a Graeco Latin square.

Manipulation and interaction. The learning environment consisted of two animated pedagogical agents that engaged in a triologue while discussing critical thinking and scientific reasoning topics (e.g., random assignment, experimenter bias, control groups) with the human learner. One agent served as a tutor agent, while the other agent served as a peer student agent. The two agents guided the human learners through the process of diagnosing flaws in eight hypothetical research studies.

During the triologue, each agent delivered their respective opinions on the quality of the research study being discussed and prompted the human learners to provide their opinions. Discrepancies between the opinions presented by the agents represented the contradictory information manipulation. This contradictory information was expected to place the human learners into a state of confusion and uncertainty.

The contradictory information was presented through four experimental conditions that varied in agent agreement and information correctness. There were two conditions in which the

two agents agreed on the quality of the research study. In the *true-true* condition both agents presented correct information (control condition), while in the *false-false* condition both agents presented incorrect information. In the two remaining conditions the two agents disagreed on the quality of the research study. In the *true-false* condition, the tutor agent presented correct information whereas the student agent disagreed by presenting incorrect information. In contrast, it was the student agent who provided the correct information and the tutor agent who disagreed with incorrect information in the *false-true* condition. It should be noted that all misleading information was corrected over the course of the dialogues and learners were fully debriefed at the end of the experiment.

The excerpt in Table 2 is an example dialogue between the two agents and the human learner. This is an excerpt from the *true-false* condition, where the tutor agent (Dr. Williams) and the student agent (Chris) are discussing a flawed research study with the human learner (Bob).

The tutorial dialogue for each learning session progressed as follows. Learning sessions began with the student agent describing the research study. The human learners were then asked to read the study and began a dialogue with the two agents. The discussion of each study involved four trials. For example, in Table 2, dialogue turns five through eight represent one trial. Each trial consisted of the student agent (turn 5) and tutor agent (turn 6) asserting their respective opinions, prompting the human learner to intervene (turn 7), and obtaining the human learner's opinion (turn 8).

These events were repeated in each trial, with each trial becoming increasingly more specific about the scientific merits of the research study. The dialogue in Table 2 discusses a research study that is flawed because it does not use random assignment. In Trial 1, the learners were simply asked if they would change their behavior based on the results of the study (turns 1-

3). Trial 2 asked if human learners believed there was a problem with the methodology of the study (turns 5-8). Trials 3 and 4 (not shown in Table 2) began to more directly address the use of random assignment. In Trial 3, human learners were asked, “Do the experimenters know that the two groups were equivalent?” Trial 4 asked the most specific question: “Should the experimenters have used random assignment here?”.

Insert Table 2 About Here

Measuring confusion and results. Confusion and uncertainty were assessed using two methods. The first method involved a retrospective affect judgment protocol (Graesser et al., 2006). Videos of learners’ faces and computer screens that were recorded during the learning session were synchronized and played back to the learners. Learners provided affect ratings over the course of viewing these videos. Learners were provided with a list of affective states (anxiety, boredom, confusion, curiosity, delight, engagement/flow, frustration, surprise, and neutral) with definitions. Affect judgments occurred at 13 pre-specified points (e.g., after contradiction presentation, after forced-choice question, after learner response) in each learning session (104 in all). In addition to these pre-specified points, learners were able to manually pause the videos and provide affect judgments at any time. Larger average values are indicative of higher levels of confusion.

Analyses with paired-samples *t*-tests of learners’ self-reported confusion revealed that the *true-false* condition ($M = .06$, $SD = .10$) had a significantly higher level of confusion than the no contradiction control condition (*true-true*; $M = .04$, $SD = .06$). However, learners in the *false-true* ($M = .04$, $SD = .06$) and *false-false* conditions ($M = .05$, $SD = .08$) did not report higher levels of confusion than the control condition. Although this finding suggests that confusion can be

induced through contradictory information, it also suggests that the degree of confusion induced is impacted by who is delivering the false information (i.e., tutor vs. student agent).

During the retrospective affect judgment protocol learners had the option to self-report the presence of eight emotions (anxiety, boredom, curiosity, delight, engagement/flow, frustration, surprise) in addition to confusion. Learners' self-reported experiences of these emotions were analyzed. When compared to the no contradiction control condition (*true-true*), learners did not self-report higher levels of any of these emotions in the experimental conditions (*true-false*, *false-true*, *false-false*). There was one exception to this finding. Curiosity was reported at a higher level in the *true-false* condition ($M = .08$, $SD = .12$) than in the *true-true* condition ($M = .04$, $SD = .07$), although the difference was only marginally significant ($p = .09$).

Self-reports are one viable method to track confusion. However, this measure is limited by learners' sensitivity and willingness to report their confusion levels. A more subtle and promising measure of confusion is to assess learner responses to forced-choice questions following contradictions by the animated agents (see turns 3 and 8 in Table 2). Since these questions adopted a two-alternative multiple-choice format, random guessing would yield a score of 0.5. Lower scores on the forced-choice questions indicate greater levels of uncertainty.

One-sample *t*-tests comparing learner responses to a chance value of 0.5 revealed the following pattern of performance: (a) *true-true* ($M = .76$, $SD = .19$) and *true-false* ($M = .60$, $SD = .19$) conditions were significantly greater than chance, (b) *false-true* ($M = .45$, $SD = .26$) was statistically indistinguishable from chance, and (c) *false-false* ($M = .35$, $SD = .31$) was significantly lower than chance. An ANOVA revealed the following pattern of response correctness across conditions: *true-true* > *true-false* > *false-true* > *false-false*.

These results suggest that contradictions successfully evoked confusion because incorrect responses following contradictions are assumed to reflect greater levels of uncertainty. The magnitude of confusion was dependent upon the source and severity of the contradiction. Confusion was low when both agents are correct and there is no contradiction (*true-true*), but increased when one agent was incorrect. Confusion was greater when the tutor agent was incorrect (*false-true*) compared to when the tutor agent was correct (*true-false*), presumably because this challenges conventional norms. Finally, confusion was greatest when both agents were incorrect (*false-false*), even though there was no contradiction in this condition.

Measuring learning and results. Learning was assessed through multiple-choice tests administered before and after the learning interaction. The multiple-choice tests consisted of 24 questions, three questions on each of the eight critical thinking concepts discussed during the learning interaction. In general, the results indicated that confusion was maximized when learners detected a clash between their knowledge and the agents' responses. It was hypothesized that the resultant confusion would provide an opportunity for deeper learning because learners would need to stop and think in order to resolve their confusion. Indeed, preliminary analyses with paired-samples *t*-tests revealed that learning was higher in the *false-true* condition ($M = .24$, $SD = .60$) than in the no contradiction control condition (*true-true*; $M = .00$, $SD = .64$). Thus, inducing confusion not only created opportunities for learning, it also resulted in increased learning, despite the fact that there was not any explicit intervention to help learners regulate their confusion.

3.4 False System Feedback during the Learning of Critical Thinking Skills

The fourth learning environment attempted to experimentally induce confusion through false system feedback. Similar to the contradictory information learning environment (Section

3.3), learners attempted to learn critical thinking and scientific reasoning skills by diagnosing flaws in research studies.

Participants and design. The participants in this experiment were 167 undergraduate students from a large mid-South university in the United States. A within-subjects design was used, such that all participants completed four learning sessions on four scientific reasoning topics. They received false (inaccurate) feedback with respect to their performance on two of the topics. Accurate feedback was provided for the remaining two topics. Order of presentation of topics and assignment of topics to condition were counterbalanced across participants using a Graeco Latin square.

Manipulation and interaction. Similar to the previous experiment (see Section 3.3), learners diagnosed flaws in research studies in an attempt to learn critical thinking and scientific reasoning skills. However, in this experiment, learners interacted with one animated pedagogical agent (tutor agent) that provided feedback on their performance. The feedback provided by the tutor agent served as the medium for inducing confusion.

The tutorial dialogue for each research study progressed as follows. Learners read a brief description of the research study and were provided with the definition of the critical thinking concept being discussed (e.g., random assignment). After receiving this information, learners were asked to determine whether the use of the critical thinking concept was flawed in the research study. After responding, learners received feedback from the tutor agent. They were then asked to read a short text on the critical thinking flaw that was present in the research study.

The quality of the learner's answer combined with the feedback delivered by the tutor agent resulted in four conditions. In the *correct-positive* condition, learners responded correctly and were given accurate, positive feedback (e.g., *Yes, that's correct*). In the *incorrect-negative*

condition, learners responded incorrectly and received negative feedback (e.g., *No, that's not correct*). These two conditions served as the control conditions. In the *correct-negative* condition, learners responded correctly but were given inaccurate, negative feedback. In the other false feedback condition, *incorrect-positive*, learners responded incorrectly but were given positive feedback. Learners receiving false feedback (*correct-negative*, *incorrect-positive*) were expected to experience more confusion than the control conditions. False feedback was expected to induce confusion in learners because it violated learners' expectations about their response quality. The false feedback was eventually corrected over the course of the dialogue.

The experiment progressed as follows. Learners first completed a pretest and read a short text introducing them to the critical thinking concepts that would be discussed during the tutorial session (control group, experimenter bias, random assignment, and replication). Learners then diagnosed flaws in four research studies, with the tutor agent providing background information, questions, and feedback. During the tutorial session learners were also provided with an explanatory text on each critical thinking concept.

Measuring confusion and results. Confusion was measured using one online method in this experiment. This method required learners to self-report the presence or absence of confusion immediately after the feedback manipulation. Although self-reported confusion was a binary choice (0 = not confused, 1 = confused) for individual learners, larger average values for a condition indicate higher levels of confusion. Mixed effects logistic regressions on these binary confusion ratings indicated that learners in the *correct-negative* condition ($M = .56, SD = .46$) reported being more confused than learners in the *correct-positive* condition ($M = .38, SD = .44$). However, there were no differences between the *incorrect-positive* condition ($M = .43, SD = .47$) and the *incorrect-negative* condition ($M = .49, SD = .47$). These findings suggest that false

system feedback can be used as a method for inducing confusion, but it is not effective in all conditions.

The findings from learners' self-reported confusion suggest that false feedback was more effective when the learner was correct (*correct-negative*) than when the learner was incorrect (*incorrect-positive*). This difference can be attributed to the difference between learners' expected outcome and the feedback received from the tutor agent. Prior to receiving feedback, learners were required to indicate whether they were confident or not confident in the correctness of their response. Regardless of actual response quality, 80% of learners reported confidence in their response correctness. This suggests that most learners believed their responses were correct, and expected to receive positive feedback from the agent. When the learners in the *correct-negative* condition received negative feedback, their expectations were violated. However, learners in the *incorrect-positive* condition had their expectations confirmed by the false feedback, and did not report increased confusion.

Learning measures and results. Learning was assessed through near and far transfer tests. Transfer tests involved learners diagnosing flaws in research studies after completing the tutorial interaction and without any guidance from the tutor agent. Each near transfer and far transfer research study contained one flaw (e.g., random assignment) that was also discussed during the learning sessions. Near transfer studies differed from the studies discussed during the learning sessions only on surface features. Far transfer studies, on the other hand, differed on both surface and structural features.

First, learning differences were investigated based on feedback condition through two mixed effects logistic regressions. The logistic regressions revealed that there were not significant differences in learners' ability to detect flaws based on feedback condition (*correct-*

negative = correct-positive and *incorrect-positive = incorrect-negative*). This finding was consistent for both near and far transfer tests.

It is possible that learners must consciously recognize the impasse in order to begin engaging in beneficial cognitive activities (VanLehn et al. 2003). To address this possibility, differences in learning outcomes based on self-reported confusion were investigated. This was investigated by testing the interaction of condition (accurate or inaccurate feedback) and self-reported confusion in two mixed effects logistic regressions (*correct-negative* vs. *correct-positive* and *incorrect-positive* vs. *incorrect-negative*).

The results indicated that both interactions were significant ($p < .05$) for far transfer tests, but the interactions were not significant for near transfer tests. Specifically, learners in both experimental conditions performed significantly better on the far transfer tests than the control conditions when confusion was reported: *correct-negative* ($M = .27, SD = .43$) performed better than *correct-positive* ($M = .12, SD = .32$) and *incorrect-positive* ($M = .41, SD = .49$) performed better than *incorrect-negative* ($M = .25, SD = .43$). When confusion was not reported learners in the *correct-negative* condition ($M = .11, SD = .30$) performed significantly worse than those in the *correct-positive* condition ($M = .26, SD = .43$), while learners in the *incorrect-positive* condition ($M = .24, SD = .41$) performed equally as well as learners in the *incorrect-negative* condition ($M = .31, SD = .46$).

To summarize, false feedback was an effective manipulation to induce confusion and promote learning. Learners who were confused by the false feedback in the experimental conditions were significantly more likely to identify flaws on the far transfer tests. This pattern of results suggests that if learners do not detect that they have reached an impasse, the learning opportunity may be missed.

3.5 Discussion

These four studies provide evidence that confusion can be induced through vague hints and prompts, breakdown scenarios, contradictions, and false feedback. Importantly, confusion was elicited through events that were tied to the learning process, as opposed to emotion induction through tasks that are independent of the learning session, such as having participants watch an emotion-inducing video (Rottenberg, Ray, & Gross, 2007). This is particularly important for implementation in learning environments that aim to regulate the induced confusion.

Also important for future ITS applications are the different methods of tracking confusion during learning. Retrospective self-reports, online self-reports, and learner responses to probing questions have been discussed as methods for tracking confusion. While there is no gold standard for detecting confusion, a combination of subjective and more objective measures of confusion was used in the present research. Specifically, utilizing both online self-reports (sporadically) and more indirect measures, such as learner responses and perhaps response times, would be the most defensible position when attempting to diagnose whether a learner is confused.

The results from the present experiments confirmed some of the predictions of goal-appraisal theory, cognitive disequilibrium theory, and impasse-driven theories of learning. In particular, the four learning environments were able to induce confusion and cognitive disequilibrium through events tied to the learning process. There was also evidence that learning opportunities were created and, in some cases, learning gains were higher in experimental conditions compared to control conditions.

4. Interventions to Help Learners Regulate their Confusion

The next step of this research program is to develop interventions to help learners regulate their confusion so as to be productively, instead of hopelessly, confused. Goleman (1995) claims, although with precious little data at hand, that expert teachers are very adept at recognizing and addressing the emotional state of learners and taking some action that positively impacts learning. But it is still an open question what these expert teachers see and how they decide on a course of action. Therefore, future research will need to empirically investigate the interventions proposed by different learning theories.

In general, a learning environment that detects that a learner is confused has a variety of paths to pursue. The learning environment might want to keep the learner confused (i.e., in a state of cognitive disequilibrium) and leave it to the learner to actively deliberate and reflect on how to restore equilibrium. This view is consistent with a Piagetian theory (1952) that stipulates that learners need to experience cognitive disequilibrium for a sufficient amount of time before they adequately deliberate and reflect via self-regulation. If so, the environment should give indirect hints and generic pumps (“okay”, “what else?”) to get the learner to do the talking when the learner flounders.

VanLehn et al. (2003) propose a method that is in line with Piagetian theory in which the ITS would only provide minor guidance while learners engage in effortful problem solving. Minor guidance from the ITS would allow learners to struggle, at times, when working to resolve their impasses. Specifically, VanLehn et al. suggest a three step strategy: “(a) let the student reach an impasse, (b) prompt them to find the right step and explain it, and (c) provide an explanation only if they have tried and failed to provide their own explanation” (pp. 245). If the ITS does provide an explanation, VanLehn et al. recommend that it should be as short and

simple as possible, excluding any extraneous information or elaboration. Alternatively, a Vygotskian theory (Vygotsky, 1978) suggests that it is not productive to have low ability students spend a long time experiencing negative affect in the face of failure. If so, the environment should give more direct hints and explanations, perhaps choosing a strategy based on the learner's prior knowledge or current knowledge level.

Individual differences, like prior knowledge, are expected to play an important role in the regulation of confusion. For example, when investigating confusion induction through device breakdowns, D'Mello and Graesser (in review) found that successful confusion resolution varied based on learners' scholastic aptitude. In these studies learners worked towards confusion resolution without any scaffolding intervention, but it is reasonable to presume that interventions would be differentially effective for individual learners. There are undoubtedly many more learner characteristics that will impact confusion experience, regulation, and ultimately the amount of learning that occurs.

For example, attribution theory (Heider, 1958; Weiner, 1995) predicts that a learner's detection of a contradiction or anomaly could be attributed to either an external cause (e.g., a problem in the learning material) or an internal cause (e.g., lack of learner knowledge). Such attributions should have an impact on the manner in which learners attempt to alleviate their confusion.

Individual differences in self-efficacy (Bandura, 1997; 2006; Zimmermann, 2000) are expected to play a role in regulation of confusion as well. One hypothesis predicts that learners with high self-efficacy will regulate their confusion through effortful problem solving and cognitive deliberation; learners with low self-efficacy, on the other hand, will experience feelings of hopelessness and disengage.

Similar to self-efficacy, learners' interest is expected to impact both the experience and regulation of confusion during learning. Learners that already possess an interest in a topic are expected to show greater perseverance when they reach an impasse, whereas learners that do not have interest in a topic are likely to disengage when faced with a challenge (Hidi & Renninger, 2006). However, it is also possible that a learning environment can create or increase learner interest in a topic.

The four phase model of interest development posits that first situational interest is triggered by an event or stimulus (Hidi & Renninger, 2006). For example, if a person read an interesting headline on a news website and read the associated article, interest in the topic could be triggered and potentially lead to seeking out more information about the topic. Then, if the correct circumstances arise, well-developed individual interest is created. When situational interest is triggered, learners will experience focused attention as well as positive feelings toward the topic that has sparked their interest. Learners who have transitioned into maintained situational interest will begin to seek out more information on the topic of interest on their own. This effort to learn more about a certain topic is intrinsically motivated and involves focused attention over a more extended period of time. Learning environments that trigger and maintain situational interest can attempt to capitalize on the expected benefits of learner interest.

The degree to which learners are academic risk takers presumably will influence the regulation of confusion. Academic risk theory contrasts (a) adventuresome learners who want to be challenged with difficult tasks, take risks of failure, and manage negative emotions when they occur, with (b) cautious learners who tackle easier tasks, take fewer risks, and minimize failure and its resulting negative emotions (Clifford, 1988; 1991; Meyer & Turner, 2006). Cautious

learners might detect the impasses but be hesitant to step up and confront their confusion because of self-doubt and the threat of failure.

In addition to pedagogical strategies that are sensitive to individual differences, another potential method of confusion regulation that would not involve complex system-learner interactions draws from emotion regulation theories (Gyurak, Gross, & Etkin, 2011; McRae et al., 2010; Sheppes & Gross, 2011). Many learners feel that being in a state of confusion is detrimental or indicative of lacking knowledge or low ability. Interventions that help learners understand the benefits of confusion, and that confusion does not imply current or eventual failure, could cause learners to be more persistent when faced with challenges. In other words, reframing the experience of confusion as positive for learning could help learners engage in the process of impasse resolution. This shift from viewing confusion as an undesirable state to a desirable state is an example of a cognitive change or cognitive reappraisal (Gross, 1998).

Future research will need to test the utility of each method and the circumstances under which each is the most useful. Within a controlled learning environment it will be possible to determine if one intervention is superior to the other methods or if there are circumstances where each is effective in producing learning gains.

5. Closing Comments

Over the last decade there has been an explosion of research on emotions during learning, both in classroom contexts and in the context of computer learning environments (see edited volumes by Schutz and Pekrun (2007) and Calvo and D'Mello (2011)). These research efforts have identified the emotions that occur during learning as well as the events that trigger these emotions and how they manifest. More recently, affect-sensitive intelligent tutoring systems that allow for the

automated management of learners' cognitive and affective states are coming online (Arroyo, et al., 2009; Burleson & Picard, 2007; Chaffar et al., 2009; Conati & Maclaren, 2009; D'Mello et al., 2009; Forbes-Riley & Litman, 2009; Robison et al., 2009). These systems are capable of automatically sensing, for example, when a learner is interested, bored, confused, and frustrated by monitoring facial cues, paralinguistic features of speech, posture, peripheral physiology, and contextual cues. Systems that couple diagnostic assessments of emotions (from sensors) with predictive assessments (from context and appraisal models) represent significant progress in this area. Emerging experimental evidence suggests that affect-sensitivity can be quite effective in promoting learning gains, especially for struggling students (D'Mello, Lehman, Sullins et al., 2010). Even more impressive are recent efforts to deploy these systems in real classrooms (Arroyo et al., 2009); this is a crucial goal for widespread use and acceptance.

Although these affect-sensitive ITSs focus on detecting certain negative emotions when they occur, research presented here suggests that there are also some merits to inducing affective states such as confusion. Based on the hypothesis that confusion can potentially be beneficial for learning, different methods for confusion induction and different methods for tracking confusion have been presented. The next step is to develop interventions that help learners productively manage their confusion so that misconceptions are corrected and conceptual networks are strengthened.

Of course it is important to emphasize that the claim is not that confusion itself is beneficial for learning or that all instances of confusion correlate with learning. Quite different from this simplistic view, the present claim is that confusion can mediate or causally effect learning if induced in the correct context and if appropriately modeled and scaffolded in a manner that is sensitive to learners' individual differences. This remains an open issue for future

research.

Acknowledgements

We thank our research colleagues in the Emotive Computing Group and the Tutoring Research Group (TRG) at the University of Memphis (<http://emotion.autotutor.org>). Special thanks to Rebekah Combs, Rosaire Daigle, Ally Dobbins, Nia Dowell, Melissa Gross, Caitlin Mills, Lydia Perkins, Amber Chauncey Strain, and Kimberly Vogt for data collection.

This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1108845) and Institute of Education Sciences (IES), U.S. Department of Education (DoE), through Grant R305A080594. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF, IES, or DoE.

References

- Ahmed, W., van der Werf, G., Minnaert, A., & Kuyper, H. (2010). Students' daily emotions in the classroom: Intra-individual variability and appraisal correlates. *British Journal of Educational Psychology, 80*, 583-597.
- Anderson, T. (2004). Toward a theory of online learning. In T. Anderson & F. Elloumi (Eds.), *Theory and practice of online learning* (pp. 33-60). Alberta, Canada: Creative Commons.
- Arroyo, I., Woolf, B., Cooper, D., Burleson, W., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In V. Dimitrova, R. Mizoguchi, B. Du Boulay & A. Graesser (Eds.), *Proceedings of 14th International Conference on Artificial Intelligence in Education* (pp. 17-24). Amsterdam: IOS Press.
- Baker, R. S., D'Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*, 223-241.
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*, 164-180.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58*, 373-403.
- Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: Introducing desirable difficulties. *American Psychological Society Observer, 19*, 3.
- Bower, G. H. (1981). Emotional mood and memory. *American Psychologist, 36*, 129-148.

- Brown, J., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.
- Burleson, W., & Picard, R. (2007). Evidence for gender specific approaches to the development of emotionally intelligent learning companions. *IEEE Intelligent Systems*, 22(4), 62-69.
- Calvo, R., & D'Mello, S. K. (Eds.). (2011). *New perspectives on affect and learning technologies*. New York, NY: Springer.
- Caroll, J., & Kay, D. (1988). Prompting, feedback, and error correction in the design of a scenario machine. *International Journal of Man-Machine Studies*, 28(1), 11-27.
- Chaffar, S., Derbali, L., & Frasson, C. (2009). Inducing positive emotional state in intelligent tutoring systems. In V. Dimitrova, R. Mizoguchi, B. Du Boulay & A. Graesser (Eds.), *Proceedings of 14th International Conference on Artificial Intelligence in Education* (pp. 716-718). Amsterdam: IOS Press.
- Chi, M.T.H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. N. Giere (Ed.), *Cognitive models of science* (pp. 129-186). Minneapolis, MN: University of Minnesota Press.
- Chi, M.T.H., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 371-399). New York, NY: Cambridge University Press.
- Clifford, M. M. (1991). Risk taking: Theoretical, empirical, and educational considerations. *Educational Psychologist*, 26, 263-298.
- Clifford, M. M. (1988). Failure tolerance and academic risk-taking in ten- to twelve-year-old students. *British Journal of Educational Psychology*, 58, 15-27.

- Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393-399.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.
- Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267-303.
- Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29*, 241-250.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper Row.
- D'Mello, S., Craig, S., Fike, K., & Graesser, A. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In J. Jacko (Ed.), *Human-computer interaction. Ambient, ubiquitous and intelligent interaction* (pp. 595-604). Berlin/Heidelberg: Springer.
- D'Mello, S. K., Craig, S. D., Witherspoon, A. W., McDaniel, B. T., and Graesser, A. C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1-2), 45-80.
- D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education, 16*, 3-28.
- D'Mello, S. K., & Graesser, A. C. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion, 25*(7), 1299-1308.

- D'Mello, S. K. & Graesser, A. C. (in press). Dynamics of affective states during complex learning. *Learning and Instruction*.
- D'Mello, S. K., & Graesser, A. C. (in review). Inducing and tracking confusion and cognitive disequilibrium with breakdown scenarios.
- D'Mello, S. K., Lehman, B. A., & Person, N. K. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, 20(4), 361-389.
- D'Mello, S. K., Lehman, B. A., Sullins, J., Daigle, R., Combs, R., Vogt, K., et al. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In J. Kay & V. Aleven (Eds.), *Proceedings of 10th International Conference on Intelligent Tutoring Systems*. Berlin/Verlag: Springer.
- Dagleish, T., Dunn, B., & Mobbs, D. (2009). Affective neuroscience: Past, present, and future. *Emotion Review*, 1(4), 355-368.
- Dagleish, T., & Power, M. J. (1999) (Eds.). *Handbook of cognition and emotion*. Chichester, NJ: Wiley.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Gosset/Putnam.
- Dettmers, S., Trautwein, U., Ludtke, O., Goetz, T. Frenzel, A. C., & Pekrun, R. (2011). Students' emotions during homework in mathematics: Testing a theoretical model of antecedents and achievement outcomes. *Contemporary Educational Psychology*, 36, 25-35.
- Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expressions: A century of research in review* (pp. 169-222). New York, NY: Academic Press.

- Ekman, P., & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.
- Forbes-Riley, K., & Litman, D. (2009). Adapting to student uncertainty improves tutoring dialogues. In V. Dimitrova, R. Mizoguchi & B. Du Boulay (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 33-40). Amsterdam: IOS Press.
- Forgas, J. P. (1995). Mood and judgment- The Affect Infusion Model (AIM). *Psychological Bulletin*, 117(1), 39-66.
- Goleman, D. (1995). *Emotional intelligence*. New York, NY: Bantam Books.
- Graesser, A. C., Chipman, P., King, B., McDaniel, B., & D'Mello, S. K. (2007). Emotions and learning with AutoTutor. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *13th International Conference on Artificial Intelligence in Education* (pp. 569-571). Amsterdam: IOS Press.
- Graesser, A. C., Lu, S. L., Jackson, G., Mitchell, H., Ventura, M., Olney, A., et al. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180-193.
- Graesser, A. C., Lu, S. L., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices breakdown. *Memory & Cognition*, 33(7), 1235-1247.
- Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95(3), 524-536.

- Graesser, A. C., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. McKeown & G. Kucan (Eds.), *Bringing reading research to life* (pp. 112-141). New York, NY: Guilford.
- Graesser, A. C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed initiative dialogue in natural language. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 243-262). Mahwah, NJ: Lawrence Erlbaum.
- Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring sessions. *Applied Cognitive Psychology, 9*, 1-28.
- Graesser, A., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine, 22*(4), 39-51. Retrieved from <Go to ISI>://000180288000004
- Graesser, A. C., Witherspoon, A., McDaniel, B., D'Mello, S. K., Chipman P., & Gholson, B. (2006). Detection of emotions during learning with AutoTutor. In R. Son (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 285-290). Mahwah, NJ: Erlbaum.
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271-299.
- Gyurak, A., Gross, J. J., & Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and Emotion, 25*(3), 400-412.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.
- Immordino-Yang, M. H., & Damasio, A. R. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain and Education, 1*(1), 3-10.

- Isen, A. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. Haviland-Jones & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548-573). New York, NY: Guilford.
- Koedinger, K., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61-78). New York, NY: Cambridge University Press.
- Kopp, K., Britt, A., Millis, K., & Graesser, A. C. (in press). Improving the efficiency of dialogue in tutoring. *Learning and Instruction*.
- Lazarus, R. S. (1999). The cognition-emotion debate: A bit of history. In T. Dagleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 3-19). Chichester, NJ: Wiley.
- Lazarus, R. (2000). The cognition-emotion debate: A bit of history. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed., pp. 1-20). New York, NY: Guilford Press.
- Lehman, B. A., D'Mello, S. K., & Person, N. K. (2010). The intricate dance between cognition and emotion during expert tutoring. In J. Kay & V. Aleven (Eds.), *Proceedings of 10th International Conference on Intelligent Tutoring Systems* (pp. 433-442). Berlin, Heidelberg: Springer.
- Lehman, B. A., D'Mello, S. K., Strain, A., Gross, M., Dobbins, A., Wallace, P., Millis, K., & Graesser, A. C. (2011). Inducing and tracking confusion with contradictions during critical thinking and scientific reasoning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of 15th International Conference on Artificial Intelligence in Education* (pp. 171-178). Berlin, Heidelberg: Springer-Verlag.
- Lehman, B. A., Matthews, M., D'Mello, S. K., & Person, N. K. (2008). What are you feeling? Investigating student affective states during expert human tutoring sessions. In B. Woolf,

- E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp. 50-59). Berlin, Heidelberg: Springer-Verlag.
- Lepper, M., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (in press). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*.
- Macaulay, D. (1988). *The way things work*. Boston: Houghton Mifflin.
- Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York, NY: Norton.
- McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society*, 13(1), 40-53.
- McRae, K., Hughes, B., Chopra, S., Gabrieli, J. D. E., Gross, J. J., & Ochsner, K. N. (2010). The neural bases of distraction and reappraisal. *Journal of Cognitive Neuroscience*, 22(2), 248-262.
- Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18(4), 377-390.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, MA: Cambridge University Press.
- Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In R. Calvo & S. D'Mello (Eds.), *New perspectives on affect and learning technologies* (pp. 23-39). New York, NY: Springer.

- Pekrun, R. (2010). Academic emotions. In T. Urdan (Ed.), *APA educational psychology handbook* (Vol. 2). Washington, DC: American Psychological Association.
- Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R. H., & Raymond, P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531-549.
- Piaget, J. (1952). *The origins of intelligence*. New York, NY: International University Press.
- Pour, P. A., Hussein, S., Al Zoubi, O., D’Mello, S. K., & Calvo, R. (2010). The impact of system feedback on learners’ affective and physiological states. In J. Kay & V. Aleven (Eds.), *Proceedings of 10th International Conference on Intelligent Tutoring Systems* (pp. 264-273). Berlin, Heidelberg: Springer.
- Psotka, J., Massey, D., & Mutter, S. (1988). *Intelligent tutoring systems: Lessons learned*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Robison, J., McQuiggan, S., & Lester, J. (2009). *Evaluating the consequences of affective feedback in intelligent tutoring systems*. In J. Cohn, A. Nijholt, & M. Pantic (Eds.), *Proceedings of the International Conference on Affective Computing & Intelligent Interaction* (pp. 37-42). Amsterdam: IEEE.
- Rodrigo, M. M. T., & Baker, R. S. J. d. (2011a). Comparing the incidence and persistence of learners’ affect during interactions with different educational software packages. In R. Calvo & S. D’Mello (Eds.), *New perspective on affect and learning technologies* (pp. 183-200). New York, NY: Springer.
- Rodrigo, M. M. T., & Baker, R. S. J. d. (2011b). Comparing learners’ affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning, 6*(1), 43-66.

- Rottenberg, J., Ray, R. D., & Gross, J. J. (2007). Emotion elicitation using films. In J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 9–28). New York, NY: Oxford University Press.
- Rus, V., & Graesser, A. C. (2007). Lexico-syntactic subsumption for textual entailment. In N. Nicolov, K. Bontcheva, G. Angelova & R. Mitkov (Eds.), *Recent advances in natural language processing IV: Selected papers from RANLP 2005* (pp. 187-196). Amsterdam: John Benjamins Publishing Company.
- Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods and research*. New York, NY: Oxford University Press.
- Schutz, P., & Pekrun, R. (2007). *Emotion in education*. San Diego, CA: Academic Press.
- Sheppes, G., & Gross, J. J. (2011). Is timing everything? Temporal considerations in emotion regulation. *Personality and Social Psychology Review*, 22, 1391-1396.
- Siegler, R., & Jenkins, E. (Eds.). (1989). *Strategy discovery and strategy generalization*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Smith, C. A., & Kirby, L. D. (2009). Relational antecedents of appraised problem-focused coping potential and its associated emotions. *Cognition and Emotion*, 23(3), 481-503.
- Stein, N. L., Hernandez, M., & Trabasso, T. (2008). Advances in modeling emotions and thought: The importance of developmental, online, and multilevel analysis. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 574-586). New York, NY: Guilford Press.
- Stein, N. L., & Levine, J. L. (1991). Making sense out of emotion: The representation and use of goal-structured knowledge. In W. Kessen, A. Ortony, & F. Craik (Eds.), *Memories*,

- thoughts, and emotions: Essays in honor of George Mandler* (pp. 295-322). England: Lawrence Erlbaum Associates, Inc.
- VanLehn, K., Graesser, A. C., Jackson, G., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31(1), 3-62.
- VanLehn, K., Siler, S., Murray, C., Yamuachi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209-249.
- Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (pp. 79-91). Cambridge, MA: Harvard University Press.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY: Guilford.
- Woolf, B. (2009). *Building intelligent interactive tutors*. Burlington, MA: Morgan Kaufmann Publishers.
- Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp. 165-184). San Diego, CA: Academic Press.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82-91.

Table 1.

Dialogue excerpt from AutoTutor interaction (Graesser et al., 2004)

Turn	Speaker	Dialogue
1	AutoTutor	When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation?
2	Learner	The passenger experiences a backward acceleration, while the car moves forward.
3	AutoTutor	And can you add to that?
4	Learner	This is because the passenger experiences a force opposite in direction to the force on the car.
5	AutoTutor	Um, yes, correct! When the neck provides a large force, what within the neck would get damaged?
6	Learner	The spine.
7	AutoTutor	Umm, I think you get this. Damage within the neck will occur in the ligaments and the ...?
8	Learner	Muscles.
9	AutoTutor	Brilliant! Ligaments and the muscles. Let's keep going. Additionally, when there is no headrest present in the car, what accelerates the head of the passenger?
10	Learner	The impact from behind.
11	AutoTutor	I am sure you can get this. The head of the passenger will be accelerated by the passenger's ...?
12	Learner	Car.
13	AutoTutor	Wrong! By the passenger's neck.
14	AutoTutor	Do you perhaps have any general conceptual physics questions?
15	Learner	No.
16	AutoTutor	How would you answer my first question, having gone over what we've just talked about?
17	Learner	The passengers get neck injuries because ... [<i>81 additional words</i>].
18	AutoTutor	Great! To recap ... [<i>AutoTutor summarizes correct response to first question</i>].

Table 2.

Dialogue excerpt from the *true-false* condition

Turn	Speaker	Dialogue Move
<p>There was experiment done at a top University where students got the same grade whether they used the textbook or not. In the fall, science students were told that textbooks were optional. For the same class in the spring, students were told that reading the textbook was required. The researchers found no differences on the final exams. So there is no need to by textbooks <i><Description of study></i></p>		
1	Dr. Williams	So Chris and I talked while you were reading and Chris thinks that there wasn't anything problematic about this study, but I think there was. <i><Introduce contradiction></i>
2	Dr. Williams	How about you Bob? Would you not buy textbooks next semester based on this study? Please type buy or not buy. <i><Forced-choice question></i>
3	Bob	Not buy. <i><Response></i>
4	Dr. Williams	We are going to go over our thinking for this study before we come to any final decisions. <i><Advance dialogue></i>
5	Chris	Well, I think how the participants were put into each condition was good, so that's not a problem. <i><Assert information></i>
6	Dr. Williams	It was problematic. <i><Contradict with Chris></i>
7	Dr. Williams	Looks like we disagree. Bob, do you think there's a problem with how the participants were put into each group? Please type problem or no problem. <i><Contradiction & forced-choice question></i>
8	Bob	Problem. <i><Response></i>

Figure Captions

Figure 1. Screenshot of AutoTutor interface

Figure 1

Auto Tutor
File Edit Session Plugins Help

How does the operating system interact with the word processing program when you create a document?

The diagram shows a computer system with the following components and interactions:

- HARD DISK**: A stack of disks at the top.
- OS**: Operating System, represented by a cloud-like shape.
- WORD PROCESSING PROGRAM**: A yellow shape within the OS.
- DOCUMENT**: A yellow shape within the Word Processing Program.
- RAM**: Random Access Memory, shown as a vertical bar on the right.
- ROM**: Read-Only Memory, shown as a pink chip.
- CPU**: Central Processing Unit, shown as a blue chip at the bottom.

Arrows indicate the flow of data: from the Hard Disk to the OS, from the OS to the Word Processing Program, from the Word Processing Program to the Document, from the Document to the CPU, from the CPU to the RAM, and from the RAM back to the OS.

Log of previous responses:

Student: the operating system allows you to save new information on a document

Tutor: I see, ok.
Tutor: Can you elaborate a bit on that?

Student: yes, the operating system creates space to save the document so that it is not lost when you open another program

Enter your response here:

yes, the operating system creates space to save the document so not lost when you open another program