Monitoring Affect States During Effortful Problem Solving Activities

Sidney K. D’Mello¹, Blair Lehman¹, and Natalie Person²

¹Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
²Department of Psychology, Rhodes College, Memphis, TN, USA

Abstract. We explored the affective states that students experienced during effortful problem solving activities. We conducted a study where 41 students solved difficult analytical reasoning problems from the Law School Admission Test. Students viewed videos of their faces and screen captures and judged their emotions from a set of 14 states (basic emotions, learning-centered emotions, and neutral) at relevant points in the problem solving process (after new problem is displayed, in the midst of problem solving, after feedback is received). The results indicated that curiosity, frustration, boredom, confusion, happiness, and anxiety were the major emotions that students experienced, while contempt, anger, sadness, fear, disgust, eureka, and surprise were rare. Follow-up analyses on the temporal dynamics of the emotions, their contextual underpinnings, and relationships to problem solving outcomes supported a general characterization of the affective dimension of problem solving. Affective states differ in: (a) their probability of occurrence as regular, routine, or sporadic emotions, (b) their temporal dynamics as persistent or random emotions, (c) their characterizations as product or process related emotions, and (d) whether they were positively or negatively related to problem solving outcomes. A synthesis of our major findings, limitations, resolutions, and implications for affect-sensitive artificial learning environments are discussed.

Keywords. Affective states, emotions, problem solving, LSAT, intelligent tutoring systems

INTRODUCTION

Learning is an immersive process that is optimized when accompanied by problem solving activities. Problem solving serves as a catalyst towards deep learning because it forces students to be active participants in their learning rather than passive information receivers (Bransford, Brown, & Cocking, 2000). Deep learning involves critically examining facts and ideas, integrating new information into existing structures, and exploring relationships between ideas. This can be contrasted with shallow learning that primarily involves the accumulation of information (Entwistle, 1988; Graesser, Jeon, & Dufty, 2008; Prosser & Trigwell, 1999). The importance of problem solving in promoting deep learning has not been overlooked by developers of Intelligent Tutoring Systems (ITSs). Several ITSs have incorporated some aspects of problem solving as part of their learning activities (Aleven & Koedinger, 2002; Alevven, McLaren, Roll, & Koedinger, 2006; Anderson, 1990; Gertner & VanLehn, 2000; Lesgold, Lajoie, Bunzo, & Eggen, 1992).

The effectiveness of problem solving in promoting learning at deeper levels of comprehension can be attributed to the deployment of key cognitive and metacognitive processing. Cognitive processes such as knowledge acquisition, information association, causal reasoning, and inference
generation coupled with metacognitive processes such as planning, re-planning, and monitoring continually operate throughout the problem solving process (Anderson, Douglass, & Qin, 2005; Brown, 1987; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Cromley, Azvedo, & Olson, 2005; Efklides, 2006; Jonassen, Peck, & Wilson, 1999). Consequently, significant attention has been devoted to the cognitive and metacognitive aspects of learning and problem solving (Anderson et al., 2005; Biggs, 1996; Boekaerts, Pintrich, & Zeidner, 2000; Graesser, D'Mello, & Person, 2009; Winnie, 2001).

With the exception of anxiety, however, the role of affective processes in deep learning and problem solving has received considerably less attention (Pekrun, Goetz, Daniels, Stupnisky, & Raymond, 2010). The affective link is critical because affective processes are inextricably bound to cognitive and metacognitive processes during learning (Barrett, 2009; Brosch, Pourtois, & Sander, 2010; Frijda, 2009; Izard, 2007; Lazarus, 2000; Moors, 2009; Russell, 2003; Scherer, 2009; Schwarz, in press; Schwarz & Skurnik, 2003; Stein, Hernandez, & Trabasso, 2008). The importance of affect is further elevated in problem solving activities because solving problems in mathematics and science is inevitably accompanied by the natural steps of making mistakes and recovering from them. Students get confused when confronted by contradictions, misconceptions, and salient contrasts (D'Mello & Graesser, 2010a; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003), they get frustrated by obstacles and challenges that block goals (Dweck, 2002; Klein, Moon, & Picard, 2002), and they experience anxiety when failure is attributed to events outside of their control (Heider, 1958; Weiner, 1986). Persistent failure might eventually lead to boredom and disengagement and other problematic long term effects such as attrition and dropout (Csikszentmihalyi, 1990; Larson & Richards, 1991; Mann & Robinson, 2009; Pekrun et al., 2010). Alternatively, students experience happiness and delight when tasks are completed, eureka moments when challenges are unveiled and major discoveries are made, and flow-like states when they are so engaged in problem solving that time and fatigue disappear (Csikszentmihalyi, 1990). Simply put, the importance of emotions during problem solving cannot be underestimated because student affective states can impede or facilitate the problem solving process by modulating cognitive processes in striking ways (Clore & Huntsinger, 2007; Fielder, 2001; Isen, 2008; Schwarz, in press; Spering, Wagener, & Funke, 2005).

Although we are beginning to make progress towards understanding the role of affect during academic problem solving, several important questions still remain unanswered. These include: (a) what are the affective states that accompany problem solving activities? (b) how does affect interact with cognitive, metacognitive, and motivational processes?, (c) how do contextual factors such as task demands, interruptions, and other perturbations impact affect?, (d) how do individual differences (e.g., affective traits, motivation, aptitude) moderate affective processes?, (e) what are the critical interactions among affect, cognition, metacognition, motivation, context, and individual differences that predict problem solving performance? There is currently no model or theory that addresses this set of questions, although some important insights have emerged over the past few decades.

Theories of Affect during Problem Solving

One of the first theories linking affect to performance is the classic Yerkes-Dodson law (1908). The law proposes a curvilinear inverted U-shaped relationship between arousal and performance that varies by task demands. Performance is low when arousal is too low or too high but is optimal when arousal is just right. The critical level of arousal to facilitate optimal performance is related to task...
complexity. In order to facilitate concentration, a lower level of arousal is optimal for tasks with high cognitive demands such as solving difficult problems.

Despite its popularity, the Yerkes-Dodson law is somewhat limited because it only addresses the arousal dimension of the affective experience, while ignoring valence. However, discriminating valenced from valenceless arousal is quite important because valenceless arousal is unlikely to impact performance (Isen, Daubman, & Nowicki, 1987). For example, climbing a flight of stairs a few times does get the heart rate up, but this type of increased arousal does not impact performance (Isen, 2003). On the other hand, positive and negative affective experiences do play a role in problem solving. For example, flexibility, creative thinking, and conceptually-driven top-down processing have been linked to positive affect (Clore & Huntsinger, 2007; Fielder, 2001; Isen, 2008; Isen et al., 1987), while negative affect has been associated with a localized, narrowly-focused, and stimulus-driven bottom-up cognitive processing (Bless & Fielder, 1995; Hertel, Neuhof, Theuer, & Kerr, 2000; Schwarz, 1990, 2000).

A two dimensional valence-arousal affect model (Barrett, 2006; Russell, 2003) does serve as a useful framework to understand the impact of affective states on problem solving activities. However, there is reason to challenge the adequacy of basing an entire theory of affect on arousal and valence alone (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Schwarz & Skurnik, 2003), because the relationship between positive and negative affective states and performance outcomes is more complex, and, in some cases counterintuitive. For example, certain negative emotions, such as confusion, could have a positive effect on problem solving because they force students to stop and think in order to resolve troublesome impasses (Craig, Graesser, Sullins, & Gholson, 2004; D’Mello & Graesser, in press-b; Graesser, Chipman, King, McDaniel, & D’Mello, 2007).

Focusing on general moods (i.e., positive or negative) during problem solving also runs the risk of overlooking the ebb and flow of dynamically changing affective states. It is more informative to monitor the fluctuations from one emotion to another and how these emotional transitions impact problem solving performance than simply noting whether a learner is generally in a positive or negative mood while solving a problem (D’Mello & Graesser, 2010a; McLeod, 1988). Therefore, a model of emotion and problem solving should extend beyond a basic valence-arousal framework and address how emotions such as confusion, frustration, and anxiety naturally arise during the problem solving process.

According to goal-appraisal theories of emotion (Mandler, 1976, 1999; Stein et al., 2008; Stein & Levine, 1991), when learners are deeply engaged, they attempt to assimilate new information into existing knowledge schemas. When new or discrepant information is detected, attention shifts to the discrepant information, the autonomic nervous system increases in arousal, and the learner experiences a variety of possible affective states depending on the context, the amount of change, and whether important goals are blocked. In the case of extreme novelty, the event evokes surprise. When there is positive feedback on an action or achievement of a difficult goal, the emotion is positive, as in the case of delight or eventually contentment. In contrast, confusion and frustration occur when the discrepancy or novelty triggers an impasse that blocks an important superordinate learning goal (e.g., solving a difficult problem or understanding a complex topic) and possibly results in the students getting stuck. The students then need to actively problem solve to resolve their confusion and frustration. For example, getting stuck and not being able to move past an obstacle due to the absence of an available plan would be interpreted as intensely negative because goal attainment is obstructed. On the other hand getting stuck, being confused, but arriving at a solution via a rare eureka moment might represent a negative to positive transition (D’Mello & Graesser, 2010a).
Research Questions

The contemporary theories of emotion (Barrett, 2009; Brosch et al., 2010; Frijda, 2009; Izard, 2007; Lazarus, 2000; Moors, 2009; Russell, 2003; Scherer, 2009; Schwarz, in press; Schwarz & Skurnik, 2003; Stein et al., 2008) convey broad links between cognition and emotions, but they do not directly explain and predict the sort of emotions that occur during learning and problem solving. There is also insufficient empirical research to conceptualize a broad, yet sufficiently detailed, model of emotional dynamics during problem solving. Therefore, the main goal of this paper is to collect and analyze data at a sufficiently detailed level of resolution as a first step towards developing such a model. Specifically, we were interested in answering the following questions: (1) What are the affective states that students experience during complex problem solving activities?, (2) What are the temporal dynamics of these affective experiences?, (3) To what extent do contextual factors (e.g., feedback, response times) impact student affect?, and (4) How are the affective states linked to performance? We addressed these questions with an exploratory study that tracked student emotions while solving difficult analytical reasoning problems. The study is exploratory because although there has been some research on monitoring emotions during one-on-one tutoring sessions with ITSs (Arroyo et al., 2009; Baker, D'Mello, Rodrigo, & Graesser, 2010; Conati & Maclaren, 2009; D'Mello & Graesser, in press-a; Forbes-Riley, Rotaru, & Litman, 2008; McQuiggan, Mott, & Lester, 2008), the research on systematic investigation of emotion during non-scaffolded problem solving is sparse and scattered.

Our focus on analytical reasoning as the problem solving domain was motivated by four primary reasons. First, these problems are complex and require considerable cognitive skills such as being attentive, efficiently representing knowledge, causal reasoning, and constraint satisfaction. Second, in addition to imposing significant cognitive demands, the problems are rife with contradictions, incongruities, anomalies, and uncertainty – all factors that are expected to trigger a diverse set of emotional reactions. Third, there are important practical implications for studying analytical reasoning problems. These problems are included in standardized tests such as the Law School Admissions Test (LSAT), a test that is required for admission into law school in the United States (LSAT, 2008). Fourth, as discussed above, several ITSs include problem solving modules due to the widely acknowledged importance of problem solving activities for deep learning. Therefore, a model of students’ emotions during effortful problem solving activities can be used to inform the development of ITSs that aspire to be responsive to students’ affective states in addition to their cognitive states.

METHOD

The study involved 41 students solving analytical reasoning problems from the Law School Admissions Test (LSAT). Students provided judgments of their emotions via a retrospective affect judgment procedure (Graesser et al., 2006). We focused on a model of student emotions consisting of: (a) basic emotions (anger, disgust, fear, happiness, sadness, surprise,) (Ekman, 1992), (b) learning-centered emotions (anxiety, boredom, confusion, contempt, curiosity, eureka, frustration) (D'Mello, Olney, & Graesser, in press; Rodrigo & Baker, in press), and (c) the non-affective neutral state.
Participants

Participants were 41 traditional college-age undergraduate students from a southern college in the United States. The participants were selected from a population of students that were enrolled in a college program that offered practice testing for graduate school standardized tests. There were 26 females (63%) and 15 males (37%). Thirty-two of the 41 students (78%) were Caucasians and the remaining nine were African-Americans (22%). All of the participants indicated that they were interested in attending law school. Participants received monetary compensation of $30 for their participation.

Procedure

The experiment was divided into two phases. Students solved analytical reasoning problems for 35 minutes in Phase 1. Phase 2 involved a retrospective affect judgment procedure.

Phase 1: Problem Solving. Students signed an informed consent before beginning the experiment. They were instructed that they would be solving 28 analytical reasoning problems and that their problem solving sessions would be videotaped. Although all of the students indicated that they were interested in attending law school, we attempted to maximize each student’s incentive to do well on the problems by offering a monetary compensation. Hence, students were informed that they would be paid two dollars for each correct answer they provided. All students were paid $30 at the end of the experiment, regardless of the amount of problems answered correctly. A sample problem is presented in the Appendix. Each problem had a scenario and approximately 5-6 sub-questions pertaining to that scenario.

Students interacted with two software programs that were displayed on a Tablet PC. The left half of the screen displayed a customized program (see Figure 1) that: (A) displayed each problem scenario, (B) asked a specific question pertaining to the scenario (i.e., a sub-problem), (C) displayed a set of answer alternatives, one of which represented the correct answer, (D) allowed students to select an answer, (E) let students finalize their choice by clicking “next”, and (F) provided feedback (“Correct”, or “Incorrect”) (not shown in Figure 1). Students were not allowed to return to a problem after it had already been completed. The program also displayed the proportion of elapsed time and the proportion of completed problems via two progress bars (see bottom of Figure 1).

Effectively solving the analytical reasoning problems required a considerable amount of thought and effort including knowledge representation, drawing diagrams, taking notes, making inferences, organizing knowledge, and other related activities. Students used a software application, Windows Journal™, to take notes and draw with a stylus. The Windows Journal program was displayed on the right half of the screen (Figure 1).

In order to expand the range of emotions that students may experience while solving the LSAT problems, we experimentally manipulated the feedback the system provided to students. Feedback was manipulated so that incorrect feedback was randomly provided for 25% of the responses (i.e., providing negative feedback to correct responses and vice versa). Therefore, there were four feedback levels, (a) correct response + positive feedback, (b) incorrect response + negative feedback, (c) correct response + negative feedback, and (d) incorrect response + positive feedback. Levels (a) and (b) represent situations where accurate feedback was provided (75% of the time), while contradictory
feedback was provided for levels (c) and (d) (25% of the time). We randomly selected between cases (c) and (d) for situations in which contradictory feedback was provided.

It should be noted that the use of false feedback was not deemed to have any harmful long-term effects because both positive and negative feedback were provided and students were not exposed to any patently incorrect domain content. Furthermore, the experimental protocol was approved prior to data collection and students were fully debriefed at the end of the experiment.

Fig. 1. Annotated screen shot of problem solving environment

The experimenter left the room after demonstrating the software interfaces to the students. Students interacted with the system for approximately 35 minutes (the analytical reasoning section of the LSAT is 35 minutes long). Four streams of information were collected during the problem solving session. First, the students’ face was recorded with a commercial webcam. The recording also included speech and other audio generated during the interaction. A video of the students’ screen was recorded using a screen capture program (Camtasia Studio™). Students’ mouse movements (i.e., the movements of the stylus of the Tablet PC) were recorded at a sampling rate of 10Hz. Finally, a time stamped log file that included information on the problems the students were working on, their responses, the feedback that was provided, what buttons were clicked, and other similar features was recorded for offline analysis.
Phase 2: Retrospective Affect Judgment. Students were given a five-minute break after the problem solving phase of the experiment. Students then completed a retrospective affect judgment protocol (Afzal & Robinson, 2009; Graesser et al., 2006; Kort, Reilly, & Picard, 2001). The procedure began by synchronizing and displaying the videos of the students’ face and screen that were captured during the interaction (see Figure 2). Videos of the screen were included to facilitate the affect judgment procedure by allowing students to incorporate contextual factors of the problem solving process with their facial expressions. Students provided affect ratings over the course of viewing these videos.

Students were provided with a list of affective states (anger, anxiety, boredom, confusion, contempt, curiosity, disgust, eureka, fear, frustration, happiness, sadness, surprise, and neutral) with definitions (see Table 1). These definitions were specifically tailored to the problem solving task. They were only designed to serve as a guide because students are more likely to rely on their personal experiences associated with experiencing, expressing, and labeling these states (i.e., folklore) (Russell, 2003). Hence, students based their judgments on: (a) videos of their faces, (b) screen captures of the Tablet PC (i.e., the context), (c) recent memories of the interaction, (d) definitions of the affective states, and (e) personal experiences associated with these emotions.

Students were required to make affect judgments at predetermined points in the session that we deemed to be relevant. These included: (a) Problem Onset - seven seconds after a new problem was displayed, (b) During Solution - halfway between the presentation of the problem and the submission of the response, and (c) After Feedback - three seconds after the feedback was provided. These time parameters were selected on the basis of pilot testing with the present system and previous research involving the retrospective affect judgment protocol (Graesser et al., 2007).

The face and screen videos automatically paused at these affect judgment points where students were required to make a judgment. The system also allowed students to rewind the videos and alter
their current judgment. In addition to the three pre-specified points where students were forced to make an affect judgment (fixed judgments), students were able to manually pause the videos and provide affect judgments at any time (spontaneous judgments).

It is worth noting three important points pertaining to the present affect judgment methodology. First, this procedure was adopted because it affords monitoring students’ affective states at multiple points, with minimal task interference, and without students knowing that these states were being monitored. Second, it has been previously used with some success (Graesser et al., 2006; Rosenberg & Ekman, 1994). Third, the offline affect annotations obtained via this retrospective protocol correlate with online recordings of facial activity and gross body movements in expected directions (D'Mello & Graesser, 2010b). Although all methods have limitations, the present method appears to be a viable approach to track emotions at a relatively fine-grained temporal resolution.

### Table 1

<table>
<thead>
<tr>
<th>Affective State</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>negative affect toward material or person to an extreme degree</td>
</tr>
<tr>
<td>Anxiety</td>
<td>nervousness, anxiety, negative self-efficacy, embarrassment</td>
</tr>
<tr>
<td>Boredom</td>
<td>uninterested in the current problem</td>
</tr>
<tr>
<td>Confusion</td>
<td>poor comprehension of material, attempts to resolve erroneous belief</td>
</tr>
<tr>
<td>Contempt</td>
<td>annoyance and/or irritation with another person</td>
</tr>
<tr>
<td>Curiosity</td>
<td>desire to acquire more knowledge or learn the material more deeply</td>
</tr>
<tr>
<td>Disgust</td>
<td>annoyance and/or irritation with the material and/or their abilities</td>
</tr>
<tr>
<td>Eureka</td>
<td>sudden realization about the material, a hah! moment</td>
</tr>
<tr>
<td>Fear</td>
<td>feelings of panic and/or extreme feelings of worry</td>
</tr>
<tr>
<td>Frustration</td>
<td>difficulty with the material and an inability to fully grasp the material</td>
</tr>
<tr>
<td>Happiness</td>
<td>satisfaction with performance, feelings of pleasure about the material</td>
</tr>
<tr>
<td>Neutral</td>
<td>displays no visible affect, at a state of homeostasis</td>
</tr>
<tr>
<td>Sadness</td>
<td>feelings of melancholy, beyond negative self-efficacy</td>
</tr>
<tr>
<td>Surprise</td>
<td>genuinely does not expect an outcome or feedback</td>
</tr>
</tbody>
</table>

## RESULTS AND DISCUSSION

The analyses were designed to address the four research questions presented in the Introduction. Analysis 1 investigated the incidence of the affective states that students experienced during problem solving. Analysis 2 involved an investigation of the temporal dynamics of these states. Analysis 3 investigated associations between contextual factors and students’ emotions. The contextual factors examined were problem phase, judgment type (fixed vs. spontaneous), feedback type, problem difficulty, and response time. Finally, Analysis 4 investigated the extent to which affective states were related to problem solving outcomes. Specifically, we investigated whether student affect while solving a problem could predict how well the problem was solved.
Analysis 1. Incidence of Affective States

A frequency analysis on the self-reported affect judgments indicated that there were 772 ($M = 19, SD = 5$) affect judgments after a new problem was presented, 768 ($M = 19, SD = 5$) judgments at the midpoint of the problem solving process, and 766 ($M = 19, SD = 5$) judgments after feedback was provided. In addition to these 2,306 judgments at fixed points (i.e., when students were required to provide an affect judgment), there were 486 spontaneous judgments ($M = 12, SD = 11$). All students provided at least one spontaneous affect judgment, although there was considerable variation among the students. Taken together, there were 2,792 ($M = 68, SD = 19$) affect judgments from 41 students. Figure 3 depicts examples of some of the affective expressions of the students.

Table 2 provides statistics on the occurrence of the various affective states. A repeated measures ANOVA indicated that there were significant differences in the proportion of emotions experienced by the students, $F(13, 520) = 35.31, MSe = .007, p < .001, \eta^2 = .469$. Bonferroni posthoc tests revealed
the following pattern in the data: (Boredom = Confusion = Curiosity = Frustration) > Anger, Anxiety, Contempt, Disgust, Eureka, Fear, Sadness, Surprise (p < .05). One minor exception to this pattern was that there were no significant differences between the occurrences of boredom and anxiety. Happiness occurred more frequently than fear and sadness, less frequently than frustration, and was on par with the other emotions.

Table 2
Distribution of affective states

<table>
<thead>
<tr>
<th>Affective States</th>
<th>Frequencies</th>
<th>Proportions</th>
<th>One-sample t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>P</td>
<td>M</td>
</tr>
<tr>
<td>Routine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td>39</td>
<td>.951</td>
<td>.106</td>
</tr>
<tr>
<td>Confusion</td>
<td>36</td>
<td>.878</td>
<td>.092</td>
</tr>
<tr>
<td>Curiosity</td>
<td>33</td>
<td>.805</td>
<td>.138</td>
</tr>
<tr>
<td>Frustration</td>
<td>39</td>
<td>.951</td>
<td>.105</td>
</tr>
<tr>
<td>Sporadic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxious</td>
<td>25</td>
<td>.610</td>
<td>.042</td>
</tr>
<tr>
<td>Happiness</td>
<td>35</td>
<td>.854</td>
<td>.055</td>
</tr>
<tr>
<td>Exceptional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contempt</td>
<td>17</td>
<td>.415</td>
<td>.027</td>
</tr>
<tr>
<td>Eureka</td>
<td>21</td>
<td>.512</td>
<td>.025</td>
</tr>
<tr>
<td>Anger</td>
<td>17</td>
<td>.415</td>
<td>.022</td>
</tr>
<tr>
<td>Disgust</td>
<td>26</td>
<td>.634</td>
<td>.030</td>
</tr>
<tr>
<td>Fear</td>
<td>6</td>
<td>.146</td>
<td>.008</td>
</tr>
<tr>
<td>Sadness</td>
<td>13</td>
<td>.317</td>
<td>.012</td>
</tr>
<tr>
<td>Surprise</td>
<td>25</td>
<td>.610</td>
<td>.027</td>
</tr>
<tr>
<td>Neutral</td>
<td>41</td>
<td>1.00</td>
<td>.311</td>
</tr>
</tbody>
</table>

Notes. N = number of students that experienced the state at least once. P = proportion of students that experienced the state at least once.

A set of follow-up analyses attempted to isolate a subset of the affective states that occurred at levels greater than chance, where chance = (1 − Mneutral)/N_affect = (1 − .311)/13 = .053. One-sample t-tests comparing the proportion of each emotion to the chance level of .053 revealed the following patterns in the data: (a) boredom, confusion, curiosity, and frustration occurred at levels greater than chance, (b) anger, contempt, disgust, eureka, fear, sadness, and surprise occurred at levels less than chance, and (c) anxiety and happiness occurred at chance levels (see Table 2).

These data support a tripartite classification of the emotions that accompany effortful problem solving activities: (a) routine emotions that include boredom, confusion, curiosity, frustration, (b)
sporadic emotions such as anxiety and happiness, and (c) exceptional emotions such as anger, contempt, disgust, eureka, fear, sadness, and surprise. The four routine emotions occurred at levels greater than chance; they comprised 64% of the observations (after excluding neutral), and, on average, these four emotions were experienced by 90% of the students. The two sporadic emotions occurred at random levels, comprised 14% of the observations, and occurred in 73% of the sessions (on average). Finally, the seven exceptional emotions occurred at levels less than chance. They collectively comprised 22% of the observations and, on average, were observed in less than half (44%) of the sessions.

Ekman’s (1992) basic emotions have been at the forefront of emotion research for the last several decades. In order to determine if they are equally applicable to problem solving, we compared the occurrence of the basic emotions (anger, disgust, fear, happiness, sadness, surprise) to the learning-centered emotions (anxiety, boredom, confusion, contempt, curiosity, eureka, frustration). A paired sample t-test confirmed that the basic emotions ($M = .154, SD = .104$) occurred at significantly lower rates than the learning-centered emotions ($M = .535, SD = .180$), $t(40) = -11.761, p < .001, d = 2.59$. This finding is on par with previous studies (D’Mello, Craig, Sullins, & Graesser, 2006; Lehman, Matthews, D’Mello, & Person, 2008) that indicate that learning activities involve more complex cognitive-affective amalgamations rather than the basic emotions.

### Analysis 2. Temporal Dynamics of Affective States

Our results so far indicate that there are graded differences in the affective experiences that accompany effortful problem solving activities. The present analysis attempted to compare the persistence of the states. Persistence refers to a property by which an affective state observed at time $t$ is also observed at $t + 1$. In other words, a state ($C$) can be considered to be persistent if experiencing it at one time interval increases the likelihood of experiencing the state at the subsequent time interval i.e., ($C_t \rightarrow C_{t+1}$). On the other hand, an affective state is ephemeral if experiencing it at $t$ decreases the likelihood that it will be observed at $t + 1$. Finally, an affective state is random if observing it at $t$ is not related to the probability of its occurrence at $t + 1$.

We utilized the Likelihood metric (Equation 1) (D’Mello, Olney, & Person, in press; D’Mello, Taylor, & Graesser, 2007) in an attempt to characterize the affective states along this tripartite classification scheme. The metric quantifies the likelihood that the current affective state ($C$) influences the next affective state ($X$) after correcting for the base rate of $X$. According to this metric, if $L(C \rightarrow X) \approx 1$, we can conclude that state $X$ reliably follows state $C$ above and beyond the prior probability of state $X$. If, on the other hand $L(C \rightarrow X) \approx 0$, then $X$ follows $C$ at the chance level. Furthermore, if $L(C \rightarrow X) < 0$, then the likelihood of state $X$ following state $C$ is much lower than the base rate of $X$.

\[
L(C \rightarrow X) = \frac{P(X|C) - P(X)}{1 - P(X)}
\]  

(Equation 1)

Our immediate goal is to assess the likelihood that affective state $C$ observed at time $t$ is also observed at time $t + 1$. This can be easily accomplished by modifying the metric such that the current state ($C$) and the next state ($X$) are the same (as illustrated in Equation 2).
\[
L(C_t \rightarrow C_{t+1}) = \frac{P(C_{t+1}|C_t) - P(C_{t+1})}{1 - P(C_{t+1})}
\]  
(Equation 2)

In order to detect significant affect state persistence, we compared the likelihood of each state repeating itself to a hypothesized mean of 0 (normalized base rate) using a one-sample \( t \)-test. The results of the tests are presented in Table 3 where it appears that the data supports a two-way classification scheme (persistent and random) instead of a three-way classification scheme, as there are no instances of ephemeral states.

<table>
<thead>
<tr>
<th>Affective State</th>
<th>Descriptives (Likelihood)</th>
<th>One-sample t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger ( \rightarrow ) Anger</td>
<td>.102</td>
<td>.214</td>
</tr>
<tr>
<td>Anxious ( \rightarrow ) Anxious</td>
<td>.097</td>
<td>.216</td>
</tr>
<tr>
<td>Boredom ( \rightarrow ) Boredom</td>
<td>.134</td>
<td>.222</td>
</tr>
<tr>
<td>Confusion ( \rightarrow ) Confusion</td>
<td>.066</td>
<td>.181</td>
</tr>
<tr>
<td>Curiosity ( \rightarrow ) Curiosity</td>
<td>.049</td>
<td>.143</td>
</tr>
<tr>
<td>Disgust ( \rightarrow ) Disgust</td>
<td>.079</td>
<td>.196</td>
</tr>
<tr>
<td>Frustration ( \rightarrow ) Frustration</td>
<td>.067</td>
<td>.151</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contempt ( \rightarrow ) Contempt</td>
<td>-.009</td>
<td>.158</td>
</tr>
<tr>
<td>Eureka ( \rightarrow ) Eureka</td>
<td>.036</td>
<td>.219</td>
</tr>
<tr>
<td>Fear ( \rightarrow ) Fear</td>
<td>.185</td>
<td>.377</td>
</tr>
<tr>
<td>Happiness ( \rightarrow ) Happiness</td>
<td>.002</td>
<td>.147</td>
</tr>
<tr>
<td>Sadness ( \rightarrow ) Sadness</td>
<td>.031</td>
<td>.153</td>
</tr>
<tr>
<td>Surprise ( \rightarrow ) Surprise</td>
<td>-.006</td>
<td>.072</td>
</tr>
</tbody>
</table>

**Analysis 3. Contextual Influences on Affective States**

The importance of context in shaping affective experience cannot be ignored for at least two primary reasons (Aviezer et al., 2008; Russell, Bachorowski, & Fernandez-Dols, 2003; Stemmler, Heldmann, Pauls, & Scherer, 2001). First, examining the context surrounding an emotional expression can lead to a deeper explanation of the emotional experience. For example, confusion while solving a problem can be contrasted with confusion after receiving feedback for the solution. The first form of confusion can be attributed to being perplexed with the problem itself, while confusion after feedback is more related to the problem solving outcome. Though similar, these two forms of confusion might have distinct manifestations and differentially impact performance.

Context also plays an important role in disambiguating between different manifestations of the same emotion. As an example consider surprise, an emotion with significant arousal but with an ambiguous valence dimension. Receiving negative feedback for a correct response would probably
evoke a degree of surprise. The valence of this exemplar of surprise is likely to be negative. But surprise could also be expected when an incorrect answer yields positive feedback. However, this type of surprise would have positive valence and might spark enthusiasm and curiosity.

In summary, context is critical because it helps disambiguating between various exemplars of a prototypical emotion (Russell, 2003). For example, the two forms of confusion discussed above are different exemplars of a prototypical “confused” state. Examining confusion (i.e., the prototype) out of context (i.e., without the exemplar) is therefore quite meaningless.

We conducted five sets of analyses that investigated relationships between the underlying context and the emotions that emerged out of that context. The contextual factors were (a) problem phase, (b) judgment type, (c) feedback, (d) problem difficulty, (e) response time, and (f) session time.

**Problem Solving Phase.** Students were prompted to make affect judgments at particular points in the problem-solving process. These judgment points occurred after the onset of a new problem (P: problem onset), midway between the presentation of the problem and the submission of the response (S: during solution), and after receiving feedback (F: after feedback). A 3 × 14 (problem phase × affect) repeated measures ANOVA investigating the distribution of emotions during the three phases of the problem-solving process (see Table 4) revealed a significant problem phase × affect interaction, $F(26, 1040) = 15.21$, $MSe = .059$, $p < .001$, partial $\eta^2 = .275$. Bonferroni posthoc tests on the problem phase × emotion interaction revealed several interesting patterns pertaining to the extent to which students experienced certain emotions at different points in the problem-solving process. For simplicity, we restrict our discussion to the more prominent affective states observed. These include boredom, confusion, frustration, curiosity, happiness, and anxiety.

<p>| <strong>Table 4</strong> Distribution of affective states during the three problem solving phases |
|---------------------------------|------|------|------|------|</p>
<table>
<thead>
<tr>
<th><strong>Affect</strong></th>
<th><strong>Problem Onset (P)</strong></th>
<th><strong>During Solution (S)</strong></th>
<th><strong>After Feedback (F)</strong></th>
<th><strong>Sig. Patterns</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>.022 .039</td>
<td>.006 .018</td>
<td>.048 .095</td>
<td>F &gt; P &gt; S</td>
</tr>
<tr>
<td>Anxious</td>
<td>.042 .045</td>
<td>.043 .057</td>
<td>.022 .049</td>
<td>P &gt; F^a</td>
</tr>
<tr>
<td>Boredom</td>
<td>.106 .108</td>
<td>.115 .108</td>
<td>.089 .146</td>
<td></td>
</tr>
<tr>
<td>Confusion</td>
<td>.092 .062</td>
<td>.135 .126</td>
<td>.029 .054</td>
<td>S &gt; P &gt; F</td>
</tr>
<tr>
<td>Contempt</td>
<td>.027 .047</td>
<td>.016 .041</td>
<td>.039 .087</td>
<td></td>
</tr>
<tr>
<td>Curiosity</td>
<td>.138 .142</td>
<td>.186 .202</td>
<td>.038 .067</td>
<td>S &gt; P &gt; F</td>
</tr>
<tr>
<td>Disgust</td>
<td>.030 .037</td>
<td>.013 .029</td>
<td>.046 .074</td>
<td>P = F &gt; S</td>
</tr>
<tr>
<td>Eureka</td>
<td>.025 .039</td>
<td>.028 .049</td>
<td>.027 .057</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>.008 .028</td>
<td>.008 .027</td>
<td>.009 .032</td>
<td></td>
</tr>
<tr>
<td>Frustration</td>
<td>.105 .071</td>
<td>.071 .081</td>
<td>.146 .142</td>
<td>F &gt; P &gt; S</td>
</tr>
<tr>
<td>Happiness</td>
<td>.055 .049</td>
<td>.012 .030</td>
<td>.144 .123</td>
<td>F &gt; P &gt; S</td>
</tr>
<tr>
<td>Neutral</td>
<td>.311 .208</td>
<td>.363 .286</td>
<td>.276 .256</td>
<td>S &gt; P = F</td>
</tr>
<tr>
<td>Sadness</td>
<td>.012 .024</td>
<td>.003 .017</td>
<td>.021 .044</td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>.027 .031</td>
<td>.002 .009</td>
<td>.066 .087</td>
<td>F &gt; P &gt; S</td>
</tr>
</tbody>
</table>

*Notes.* ^a^ $P = S$ and $S = F$ for anxiety
The results indicated that boredom was evenly distributed across all three judgment points (i.e., \( P = S = F \)). This suggests that when students disengage, potential affect-inducing events such as the presentation of a new problem or feedback do little to alleviate their boredom. On the other hand, confusion and curiosity displayed a strikingly different (compared to boredom) pattern of occurrence (i.e., \( S > P > F \)). These emotions were most frequently observed in the midst of problem solving, followed by the presentation of a new problem. Experiences of confusion and curiosity were rare after feedback was provided.

Frustration and happiness were another pair of affective states with similar occurrence patterns (i.e., \( F > P > S \)). These states were most frequent after feedback was provided followed by the onset of a problem. Frustration and happiness rarely occurred during the process of deriving a solution to the problem. These affective states seem to be complimentary to confusion and curiosity in that they occur after a resolution has been reached (i.e., feedback), as opposed to confusion and curiosity that occur while the student tries to arrive at a solution. Consequently, confusion and curiosity appear to be related to the problem solving process, while frustration and happiness are related to the problem solving outcome (or product). Furthermore, although their occurrences are rare, anger and surprise also appear to be linked to the outcome of problem solving, which is what could be expected.

**Judgment Type.** In addition to being required to provide judgments at the predetermined fixed points, students also had the option of providing judgments at any time during the retrospective judgment procedure (spontaneous judgments). The fixed judgments consisted of our best guess of when an emotion was most likely to occur. However, the spontaneous judgments represented instances where our estimated judgment points missed the mark. Therefore, it is quite conceivable that a different set of affective responses can be elicited from the fixed judgment points (i.e., problem onset, during solution, and after feedback) when compared to spontaneous judgments points.

A \( 2 \times 14 \) (judgment type \( \times \) affect) repeated measures ANOVA on the distribution of emotions for fixed versus spontaneous judgments revealed a significant judgment type \( \times \) affect interaction, \( F(13, 520) = 2.29, MSe = .010, p < .01 \). However, the effect size for this interaction (partial \( \eta^2 = .054 \)) was substantially smaller than the problem phase \( \times \) affect interaction (partial \( \eta^2 = .275 \)), indicating that most of the variance was explained by the differences in emotions observed at the different points in the problem solving process.

Nevertheless, there were occasional differences in the distribution of emotions across the two judgment types. Bonferroni posthoc tests revealed that confusion was reported at significantly higher rates for the spontaneous (\( M = .164, SD = .167 \)) than the fixed (\( M = .086, SD = .065 \)) judgment points. A different pattern emerged for curiosity and happiness. Both these emotions were observed at higher rates for the fixed judgments (curiosity: \( M = .121, SD = .130 \); happiness: \( M = .071, SD = .059 \)) when compared to the spontaneous judgments (curiosity: \( M = .079, SD = .136 \); happiness: \( M = .026, SD = .069 \)).

**Feedback.** Feedback is critical in both human and computer tutoring because it is directive (i.e., tells students what needs to be fixed), facilitative (i.e., helps students conceptualize information), and has motivational functions (Black & William, 1998; Lepper & Woolverton, 2002; Shute, 2008). Feedback strategies of tutors have received considerable attention from educational researchers, with a handful of meta-analyses devoted exclusively to the effectiveness of feedback as a pedagogical and motivational tool (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Shute,
However, relatively little is known about the impact of feedback on students’ affective states, although some data is beginning to emerge (D’Mello et al., 2006; D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008).

The study included an experimental manipulation pertaining to the type of feedback the system provided to students. Feedback was provided after the students submitted an answer to a problem and can be subdivided into four categories on the basis of the quality of the student responses (correct or incorrect) and the type of feedback provided by the system (positive or negative). The four feedback categories are: (a) PP = correct response + positive feedback, (b) NN = incorrect response + negative feedback, (c) PN = correct response + negative feedback, and (d) NP = incorrect response + positive feedback. Accurate feedback was provided for categories PP and NN while contradictory feedback was provided for categories PN and NP.

We performed a $4 \times 14$ (feedback type × affect) repeated measures ANOVA to investigate associations between feedback and student affect after feedback was provided. Since our major focus is on emotions after feedback, we excluded spontaneous affect judgments as well as judgments after problem presentation and during problem solution. The results revealed that there was a significant feedback type × affect interaction, $F(39, 1560) = 8.92$, $MSe = .023$, $p < .001$, partial $\eta^2 = .182$.

Bonferroni posthoc tests comparing the occurrence of each emotion across the different feedback categories support several important conclusions (see Table 5).

Posthoc tests showed that boredom and curiosity were not related to feedback, confusion was somewhat related, and frustration and happiness were strongly related to the type of feedback provided. The fact that the occurrence of boredom was independent of the type of feedback provided supports a general characterization of boredom during learning activities. Bored students essentially disengage from the learning session to a point where external stimulation (via a new problem or feedback) does little to remove students out of a state of persistent ennui. Simply put, boredom begets more boredom.

A rather different interpretation can be applied to curiosity not being related to the type of feedback. Curiosity is more prominent during the problem solving process (discussed above). Therefore, the fact that feedback was not related to levels of curiosity strengthens the characterization of curiosity as a process-related emotion.

Similar to curiosity, earlier we characterized confusion as a process-related emotion. However, it appears that confusion is more prominent after negative versus positive feedback. Taken together, these results support a more refined categorization of confusion, because this emotion appears to have complimentary process- as well as product-related manifestations. Students experience confusion during the problem solving process and also when they are provided negative feedback to an answer they think is correct (i.e., the product).

Consistent with their characterization as product-related emotions, frustration and happiness were highly related to feedback. Frustration occurred when negative feedback was provided regardless of whether the feedback was accurate (NN) or contradictory (PN). A reverse pattern was observed for happiness. This state was more prominent when positive feedback was provided irrespective of whether the feedback was accurate (PP) or contradictory (NP). Therefore, the two product-related emotions appear to be frustration (negative) and happiness (positive).

Contradictory to our expectations, it appeared that providing contradictory feedback was not noticeably associated with students’ emotions. Instead, it is the valence (PP, NN) rather than the presence or absence of a contradiction (PN, NP) that influenced most of the feedback-related effects. It
might be the case that the students were simply unaware of the contradictory feedback. This is a plausible speculation since contradictory feedback was only provided for 25% of the cases.

**Table 5**

<table>
<thead>
<tr>
<th>Affect</th>
<th>PP M</th>
<th>PP SD</th>
<th>NN M</th>
<th>NN SD</th>
<th>PN M</th>
<th>PN SD</th>
<th>NP M</th>
<th>NP SD</th>
<th>Sig. Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>.034</td>
<td>.103</td>
<td>.095</td>
<td>.176</td>
<td>.077</td>
<td>.229</td>
<td>.000</td>
<td>.000</td>
<td>NN &gt; NP = PP</td>
</tr>
<tr>
<td>Anxiety</td>
<td>.055</td>
<td>.146</td>
<td>.011</td>
<td>.044</td>
<td>.041</td>
<td>.170</td>
<td>.037</td>
<td>.165</td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td>.005</td>
<td>.046</td>
<td>.017</td>
<td>.083</td>
<td>.024</td>
<td>.109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusion</td>
<td>.016</td>
<td>.058</td>
<td>.038</td>
<td>.137</td>
<td>.037</td>
<td>.173</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contempt</td>
<td>.061</td>
<td>.128</td>
<td>.006</td>
<td>.039</td>
<td>.000</td>
<td>.000</td>
<td>.033</td>
<td>.130</td>
<td>PP &gt; NN = PN</td>
</tr>
<tr>
<td>Curiosity</td>
<td>.005</td>
<td>.078</td>
<td>.071</td>
<td>.187</td>
<td>.012</td>
<td>.078</td>
<td>NN &gt; NP = PP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>.033</td>
<td>.022</td>
<td>.013</td>
<td>.044</td>
<td>.008</td>
<td>.052</td>
<td>.004</td>
<td>.026</td>
<td></td>
</tr>
<tr>
<td>Eureka</td>
<td>.009</td>
<td>.040</td>
<td>.014</td>
<td>.047</td>
<td>.000</td>
<td>.000</td>
<td>.174</td>
<td>.303</td>
<td>PP &gt; NP &gt; NN = PN</td>
</tr>
<tr>
<td>Fear</td>
<td>.333</td>
<td>.306</td>
<td>.249</td>
<td>.250</td>
<td>.266</td>
<td>.359</td>
<td>.060</td>
<td>.223</td>
<td>NN = PN &lt; NP = PP</td>
</tr>
<tr>
<td>Frustration</td>
<td>.381</td>
<td>.317</td>
<td>.183</td>
<td>.270</td>
<td>.131</td>
<td>.278</td>
<td>.280</td>
<td>.406</td>
<td>PP &gt; NP &gt; NN = PN</td>
</tr>
<tr>
<td>Happiness</td>
<td>.000</td>
<td>.000</td>
<td>.042</td>
<td>.094</td>
<td>.032</td>
<td>.106</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>.036</td>
<td>.088</td>
<td>.032</td>
<td>.073</td>
<td>.107</td>
<td>.196</td>
<td>.126</td>
<td>.260</td>
<td>NN &gt; NP = PP</td>
</tr>
</tbody>
</table>

**Notes**

*PN > NP for disgust*

**Problem Difficulty.** We investigated the extent to which problem difficulty was associated with students’ emotions. We focused on the emotions that occurred while the students were in the process of solving the problem (i.e., affect judgments after problem presentation, feedback, and spontaneous judgments were excluded). Our analyses proceeded by computing a difficulty value for each of the 28 problems. The difficulty level of a problem was operationalized as the proportion of students that correctly solved it. In this fashion, problem difficulty was computed for all 28 problems and was normally distributed ($M = .432, SD = .285$). We dichotomized this variable into low and high difficulty problems via a median split ($median = .432$).

A $2 \times 14$ (difficulty level $\times$ affect) repeated measures ANOVA indicated that there was a significant but small interaction between problem difficulty and affect, $F(13, 520) = 2.65, MSE = .006, p < .01$, partial $\eta^2 = .062$. Bonferroni posthoc tests revealed that the significant difference was for boredom, but not for any of the other affective states. It was found that boredom occurred at a higher rate for easy ($M = .139, SD = .151$) compared to difficult problems ($M = .086, SD = .113$). So as could be expected, problems that do not sufficiently challenge and engage students are linked to boredom, a finding which is consistent with flow theory (Csikszentmihalyi, 1990) and the control-value theory of emotions (Pekrun et al., 2010).

**Response Time.** We investigated associations between response time (i.e., time between problem presentation and answer submission) and student emotions. We focused on the emotions that occurred while the students were in the middle of solving the problem and excluded judgments after problem
presentation, feedback, and spontaneous points. Our analyses proceeded by dichotomizing each student’s response times into a low and high group via a median split procedure (median = 73.8 seconds).

A 2 × 14 (response time × affect) repeated measures ANOVA indicated that there was a significant interaction between response time and student affect, \( F(13, 520) = 3.04, MSe = .006, p < .001, \) partial \( \eta^2 = .071 \). The significant differences were for boredom and confusion but not for any of the other emotions. It appears that students have shorter response times when they are bored (\( M_{low} = .146, SD_{low} = .152, M_{high} = .083, SD_{low} = .109 \)), but a reverse pattern is observed when students are confused (\( M_{low} = .10, SD_{low} = .113, M_{high} = .172, SD_{low} = .187 \)). One interpretation to this finding is that bored students have faster response times because they tend to move ahead without investing any substantial cognitive resources in the task. Conversely, confused students have longer response times because they take the time to effortfully deliberate in order to alleviate their perplexity.

Session Time. We investigated the extent to which session time was associated with the affective states of the student. It is quite conceivable that students begin the session with more enthusiasm and curiosity, which gradually transitions into boredom as time progresses. For example, within the context of learning with ITSs, there is some evidence that confusion occurs earlier in a tutoring session, while boredom increases as the session progresses (D'Mello, Craig et al., 2008).

The analysis proceeded by dividing each 35-minute session into three 10-minute time intervals (the last 5 minutes were ignored). We then computed the distribution of emotions for each of the time intervals. A 3 × 14 (time interval × affect) ANOVA indicated that there was no significant time interval × affect interaction (\( p = .506 \)).

Analysis 4. Affect and Problem Solving Outcomes

Students were required to solve 28 difficult problems in approximately 35 minutes. However, on average they only solved 19 problems in the allotted time yielding a mean completion rate of 67.5% (\( SD = 18.8 \)). The mean precision score was .46 (\( SD = .15 \)) and the mean recall was .308 (\( SD = .133 \)).

The pertinent question of whether student affect was related to problem solving performance was addressed by comparing the distribution of emotions for problems that students answered correctly to problems that were answered incorrectly.

Fixed Judgment Points

After Feedback on Previous Problem. We were unable to investigate whether feedback on the current problem (\( R_n \)) was related to the outcome of its solution because feedback on the current

---

1 A warning indicating that time is running out was automatically issued during the last 5 minutes of the problem solving phase. After receiving this warning, participants demonstrated heightened anxiety and a general state of panic. The emotions experienced during the last 5 minutes were not included in these analyses as they are more of an artifact of the warning message rather than the natural steps of problem solving.

2 Precision = \( \frac{\# \text{ of problems correctly answered}}{\# \text{ of problems attempted}} \); Recall = \( \frac{\# \text{ of problems correctly answered}}{\text{total } \# \text{ of problems}} \)
problem is provided after its solution. However, we were able to investigate whether feedback on the previous problem \( P_{n-1} \) could predict whether the current problem was answered correctly.

A 2 \( \times \) 14 (response type [correct | incorrect] \( \times \) affect) ANOVA revealed that student affect after receiving feedback for problem \( P_{n-1} \) was related to the outcome for problem \( P_n \), \( F(13, 520) = 2.19 \), \( MSe = .008 \), \( p < .01 \), partial \( \eta^2 = .052 \). Bonferroni posthoc tests revealed that the significant differences were only for the product-related emotions of frustration and happiness. It appears that increased levels of frustration after receiving feedback for the previous problem was linked to diminished performance on the current problem, \( \text{Frustration}_{\text{CORRECT}} (M = .121, SD = .149) < \text{Frustration}_{\text{INCORRECT}} (M = .173, SD = .172) \). On the other hand, being happy after receiving feedback on the previous problem was linked to enhanced performance on the current problem, \( \text{Happiness}_{\text{CORRECT}} (M = .176, SD = .184) > \text{Happiness}_{\text{INCORRECT}} (M = .111, SD = .103) \).

**After Onset of Current Problem.** An ANOVA indicated that student affect a few seconds after a problem was presented was related to how well they solved the problem, \( F(13, 520) = 1.76 \), \( MSe = .007 \), \( p < .05 \), partial \( \eta^2 = .042 \). It appears that being curious after a problem statement was presented was related to how well the problem was solved, \( \text{Curiosity}_{\text{CORRECT}} (M = .243, SD = .238) > \text{Curiosity}_{\text{INCORRECT}} (M = .176, SD = .227) \).

**During Solution of Current Problem.** An ANOVA indicated that affective states in the midst of solving a problem was related to problem solving outcomes, \( F(13, 520) = 2.37 \), \( MSe = .006 \), \( p < .01 \), partial \( \eta^2 = .056 \). Bonferroni posthoc tests revealed that enhanced curiosity while solving a problem was linked to positive outcomes, \( \text{Curiosity}_{\text{CORRECT}} (M = .216, SD = .237) > \text{Curiosity}_{\text{INCORRECT}} (M = .161, SD = .204) \). A reverse pattern emerged for frustration. It appears that increased frustration during the problem solving process was negatively related to performance, \( \text{Frustration}_{\text{CORRECT}} (M = .05, SD = .089) < \text{Frustration}_{\text{INCORRECT}} (M = .084, SD = .095) \). Therefore, being curious immediately after being presented with the problem statement and maintaining this level of curiosity while solving the problem, is associated with positive outcomes. In contrast being frustrated after receiving feedback on the previous problem, and sustaining this negative emotion while solving the current problem, is unfavorable for performance.

As could be expected we discovered that experiences of eureka were positively linked to positive outcomes, \( \text{Eureka}_{\text{CORRECT}} (M = .04, SD = .073) > \text{Eureka}_{\text{INCORRECT}} (M = .014, SD = .043) \). Therefore, although “a-ha” moments that resonate with the eureka experience are rare, they are valuable indicators of successful outcomes.

**Spontaneously Reported Affective States**

An ANOVA confirmed that there was a significant relationship between spontaneously reported affect and problem solving outcomes, \( F(13, 364) = 3.86 \), \( MSe = .017 \), \( p < .001 \), partial \( \eta^2 = .121 \). Bonferroni posthoc tests revealed that spontaneous reports of disgust during the problem solving process were negatively linked to performance, \( \text{Disgust}_{\text{CORRECT}} (M = .002, SD = .008) < \text{Disgust}_{\text{INCORRECT}} (M = .077, SD = .135) \). There were no differences for any of the other emotions. Therefore, although disgust rarely occurs (3\% of the emotions), students that are disgusted to a point that they feel the need to voluntarily report their disgust rarely provide a correct answer.
Discussion on Relationships between Emotions on Problem Solving Outcomes

The links between students’ affective states at the various phases of the problem solving process in conjunction with spontaneously reported emotions highlight important relationships between emotions and problem solving outcomes. It appears that students anguish over past failure because frustration after receiving negative feedback on the previous problem was negatively related to the quality of the solution associated with the current problem. However, students also revel in past successes. Happiness associated with a positive outcome on the previous problem was positively linked to success on the current problem.

An example of the relationship between affect and problem solving outcomes is illustrated in Figure 4. The figure depicts moving averages of performance (dotted lines) and relevant affective states (solid lines) for a sample student.

Fig. 4. Affective states and problem solving performance for a sample student.

This student appears to oscillate between bouts of failure (problems 1-3 and 10-17) and success (problems 4-9 and 17-22). As illustrated in Figure 4A, the initial drop in performance is mirrored by an increase in frustration (problems 1-3), while frustration diminishes when problems 4-9 are correctly answered. A new round of failure (problems 10-17) is accompanied by a small rise in frustration,
which stays constant as performance gradually improves (problems 17-22). It is intriguing to note that happiness and frustration show an excitatory-inhibitory relationship as depicted in Figure 4B. Early failure that increases frustration is accompanied by a drop in happiness. But happiness increases as performance improves and frustration dissipates.

The results also indicated that being curious when a new problem is presented and maintaining that level of curiosity during the problem solving process (i.e., after problem presentation and before answer submission) was positively associated with performance. This alignment between curiosity and problem solving performance is exemplified in Figure 4C. Failure is accompanied by a drop in curiosity, but curiosity levels are rejuvenated when performance improves.

Finally, although confusion was not statistically linked to problem solving outcomes, this sample student’s experiences of confusion were closely linked to performance (Figure 4D). For example, initial poor performance is associated with a sharp increase in confusion while confusion dissipates as performance improves.

**GENERAL DISCUSSION**

Our results support the hypothesis that emotions play a critical role in effortful problem solving activities. Since emotion and motivation are inextricably bound to learning (Meyer & Turner, 2006; Snow, Corno, & Jackson, 1996), the diverse and rich emotional tapestry observed in this study provides important clues into students’ motivational levels. The students in this study were sampled from a population of students that planned on taking the LSAT in the near future, ostensibly because law school was a viable career choice. The alignment between the task (i.e., solving sample LSAT problems) and goals (i.e., taking the actual LSAT for law school admission) may have positively influenced their motivation levels. Students’ motivation levels could also have been enhanced with the monetary incentive of two dollars for each question they answered correctly. Although monetary incentives may not be common in real world learning situations, other extrinsic motivators are part and parcel of most educational tasks. Whether it is parental approval, completion of a class, acceptance into college, or obtaining a promotion, the learning environment cannot be divorced from the realities of the real world, which is rife with sources of extrinsic motivation.

Our multi-faceted investigation on the occurrence of the emotions, their temporal dynamics, contextual underpinnings of problem phase, feedback, response time, and problem difficulty, and on relationships between affect and problem solving outcomes supports some important conclusions into the affective dimension of problem solving. We proceed by taking stock of the major findings by assimilating the various analysis threads, listing some of the limitations of the study and offering potential solutions, and discussing applications of our findings for ITSs.

**Assimilation of Major Findings**

The results indicated that affective states differ in their probability of occurrence as regular, routine, or sporadic emotions, their temporal behavior as persistent or random emotions, their manifestations as product or process related emotions, and whether they facilitate or impede the problem solving process. An examination of the more frequent emotions along these dimensions provides a broad picture of the role of emotions during problem solving (see Table 6).
Curiosity and Boredom. Curiosity was the most frequent emotion and was strongly associated with problem solving performance. Curiosity can be characterized as a routine and persistent emotion that is intimately related to the problem solving process and positively predicts performance. Berlyne (1978) describes curiosity as a form of deliberate, exploratory behavior, which in motivated students is a natural correlate to solving problems that are challenging, exciting, and require an active exploration of the problem space to arrive at a solution. The fact that curiosity is related to interest (Berlyne, 1960, 1978; Deci, 1992; Izard & Ackerman, 2000; Tobias, 1994), is one possible explanation why curiosity was positively linked to performance.

Boredom, on the other hand, is the antithesis of interest, engagement, and curiosity. Boredom routinely occurs, is quite persistent, and is neither a product nor a process related emotion. Bored students disengage to an extent where any external stimulation via feedback or a new problem is ineffective in capturing their interest. This pattern of boredom is consistent with previous research that tracked boredom during tutorial sessions with ITSs (Baker et al., 2010; D’Mello & Graesser, in press-b). Although boredom was not negatively associated with performance in the present study, we suspect that this lack of a relationship might be attributed to the relatively short 35-minute problem solving session. Boredom might have a more negative effect over longer time spans. For example, there is considerable research that highlights the detrimental effects of boredom such as negligible learning, lower self-efficacy, diminished interest in educational activities, and, most importantly, increased attrition and dropout (Craig et al., 2004; Farrell, Peguero, Lindsey, & White, 1988; Larson & Richards, 1991; Mann & Robinson, 2009; Pekrun et al., 2010; Perkins & Hill, 1985; Robinson, 1975). Taken together, these results emphasize the importance alleviating boredom while promoting curiosity and interest so students might pursue more productive trajectories of thought.

Frustration and Happiness. Frustration and happiness were the two major product-related emotions. Frustration accompanied negative feedback while happiness was related to positive feedback. Since students provided correct responses to approximately 50% of the problems, one would expect the incidence of these two affective states to be approximately equal. However, frustration occurred at twice the rate as happiness, suggesting that the valence dimension alone does not provide an adequate
explanation on the occurrence of these two emotions. An examination of these emotions along a temporal dimension provides some additional insights. It appears that frustrated students are more likely to stay frustrated, while happy students routinely transition into other affective states. Simply put, frustration is unrelenting while happiness is fleeting.

**Confusion.** Confusion appears to be a routine and persistent emotion with both product- and process-related manifestations, a finding that substantiates existing research that highlights the importance of this emotion to problem solving and deep thinking (Festinger, 1957; Graesser et al., 2005; Graesser & Olde, 2003; Piaget, 1952; Rozin & Cohen, 2003; Silvia, 2009; VanLehn et al., 2003). The high levels of confusion could be attributed to the problems being riddled with complications, salient contrasts, and other obstacles – all factors that put students in a state of cognitive disequilibrium (Graesser et al., 2005; Graesser & Olde, 2003). Confusion is often accompanied by effortful cognitive activities as students try to resolve impasses, thus returning to a state of cognitive equilibrium. This can be considered to be a form of productive confusion that forces students to stop and think. But confusion also has a less productive form, namely hopeless confusion, where students are unable to achieve a resolution and get stuck. Unresolved confusion transitions into frustration which over time triggers boredom (D'Mello & Graesser, 2010a). We suspect that alternating between these two forms of confusion might explain why this emotion was not explicitly linked to problem solving outcomes.

**Limitations and Possible Resolutions**

There are four primary limitations with this study. Perhaps the most significant limitation pertains to the fact that the correlational design we adopted limits our ability to draw causal relationships between the primary variables (e.g., emotions and problem solving outcomes). It is possible to experimentally induce particular affective states so that causal links can be tested, however, most of the methods for affect induction, such as exposure to affectively charged stimuli such as films and images (Coan & Allen, 2007), are divorced from the context of the primary task (i.e., problem solving). For example, the anxiety induced from exposing participants to images of spiders (Ohman & Soares, 1993) is ostensibly quite different from anxiety associated with poor performance during problem solving. Hence, the added experimenter control afforded by context-free affect induction techniques is accompanied by a substantial drop in ecological validity. This is a compromise that is incompatible with the present goal of monitoring the antecedents and consequents of naturally occurring emotions that are inextricably bound to the problem solving process. Nevertheless, while the present paper focuses on correlational links between a large set of variables, future work should systematically induce affect in a context-sensitive fashion so that important causal links can be tested.

A second limitation pertains to exclusively relying on self-reports for affect measurement. Self-reports are limited by the participant’s ability and sensitivity to his or her emotions and the accuracy of the reports depends on the honesty of the student. Self-reports also require the emotion to be consciously accessible, so subtle unconscious affect experiences are likely to be missed. Hence, it is important to confirm that the major findings of the study are replicated when alternate methods are used to measure student emotions. Some alternate methods include live observations (Craig et al., 2004; Rodrigo & Baker, in press), offline judgments by peers, trained judges, and teachers (D'Mello, Taylor, Davidson, & Graesser, 2008; Graesser et al., 2006), and physiological and behavioral instrumentation (Arroyo et al., 2009; Calvo & D'Mello, 2010). Perhaps the most defensible position is to consider multiple measures in order to obtain more reliable and valid measurements.
The third and fourth limitation can be attributed to students being required to make forced-choice affect judgments and that mixed emotions (e.g., confusion + anxiety) were not tracked. It is important that these issues be addressed in replication studies where mixed emotions are monitored and students have some flexibility in making their judgments. For example, students could be provided with the option of selecting an “other” category, reporting combinations of emotions that they are simultaneously experiencing, or using a dimensional measurement instrument like the Affect Grid (Russell, Weiss, & Mendelsohn, 1989).

Applications of our Findings for ITSs

Tutoring in mathematics and science routinely involves periods of problem solving (Aleven & Koedinger, 2002; Aleven et al., 2006; Anderson, 1990; Gertner & VanLehn, 2000; Lesgold et al., 1992). These effortful problem solving activities often involve failure, and students experience a host of negative emotions such as frustration, anxiety, and boredom. There is no sterile learning environment that always promotes positive affective experiences because negative emotions inevitably accompany the natural steps of making mistakes, being stuck, and resolving impasses. The challenge for ITS developers is to leverage emerging basic research on affect and learning towards the development of computer environments that promote learning at deeper levels of comprehension and mastery in a manner that effectively coordinates cognition and emotions. The present study directly contributes to this goal of developing affect-sensitive ITSs that detect and respond to students’ emotional states in addition to their cognitive states.

In particular, we have identified curiosity, boredom, confusion, and frustration as the major affective states that students experienced during problem solving. Anxiety and happiness were occasionally experienced, while the remaining six emotions were rare. Importantly, with the exception of happiness, the basic emotions did not appear to be very relevant, at least within the 35-minute problem solving sessions adopted in this study. Although this finding warrants replication with diverse student populations, different problem solving contexts, and alternate methodologies to track emotions, it is consistent with previous studies that have monitored emotions during tutoring sessions with ITSs (D’Mello et al., 2006) and expert human tutors (Lehman et al., 2008). Taken together, these results are suggestive of an important point of divergence between general affective-computing research, which primarily focuses on the basic-emotions (Calvo & D’Mello, 2010; Zeng, Pantic, Roisman, & Huang, 2009), and the specialized niche of affective learning environments (Calvo & D’Mello, in preparation) where the learning-centered emotions are more prominent. It is these learning-centered emotions that should be on the radar of affect-sensitive ITSs.

While the incidence data narrows the landscape of relevant emotions, the persistence data provides some clues into the relative importance of responding to these emotions. It appears that in addition to being very frequent, boredom is also one of the most persistent emotions. Once triggered, boredom adopts a persistent temporal quality and students risk disengaging to a point that any further instruction is essentially futile. There is also some evidence that tutorial interventions are not very effective at alleviating boredom (D’Mello & Graesser, in press-b), so a proactive strategy of predicting and preventing boredom might be more productive than a reactive strategy of detection and regulation. On the other hand, curiosity, which was closely associated with productive problem solving leading to successful performance, was comparatively less persistent. This suggests that it might be important for ITSs to implement interventions that promote curiosity and engagement in order to ignite and nurture sparks of interest that can be sustained over multiple sessions.
The results from our analysis of the contextual cues (e.g., problem phase, feedback) surrounding emotional experiences can be used to scaffold the development of automatic affect detection systems. Emotion detection is a major challenge for affect-sensitive ITSs because an ITS can never respond to student emotions if it cannot detect those emotions. Although most systems rely on physiological and bodily cues to detect emotions (Arroyo et al., 2009; Castellano, Kessous, & Caridakis, 2008; D’Mello & Graesser, 2010b), the link between affective experience and expression is diffuse, fuzzy, murky, and possibly indeterminate (Calvo & D’Mello, 2010; Russell et al., 2003). Since context plays a major role in shaping emotional experiences, it is important to couple bottom-up diagnostic assessments of affective reactions (i.e., face, speech, posture) with top-down contextually-driven predictive assessments (Conati & Maclaren, 2009). For example, we know that some emotions are closely related to the problem solving process, while others are tightly coupled to the feedback received (i.e., the product). This information can be used to develop context-based affect prediction systems that guide physiological and bodily-based affect detectors.

In summary, we provided a sketch of how some of the insights obtained from the present analysis of affect during problem solving can be applied to advanced learning technologies that sense and respond to student affect. What we have not addressed is explicit strategies to regulate negative emotions and promote more productive states. Although this question is beyond the scope of the present study, it is possible to speculate on some possible affect regulation strategies that can be implemented in ITSs. For example, an ITS that senses negative emotions can respond by offering hints to diffuse confusion, empathetic statements to alleviate frustration, motivational statements to avert anxiety, and novel challenges to offset boredom (Burleson & Picard, 2007; D’Mello, Craig, Fike, & Graesser, 2009; Forbes-Riley & Litman, 2010; Robison, McQuiggan, & Lester, 2009; Woolf et al., 2010). Positive emotions such as curiosity, engagement, and interest could be fostered by encouraging students to select their own problems since these emotions can be stimulated in environments that foster students’ freedom of choice and when students perceive a degree of value in the learning activity (Guthrie & Alvermann, 1999; Lepper & Woolverton, 2002; Pekrun, 2010; Pekrun et al., 2010). It is our hope that these affect-sensitive interventions will fortify students with the necessary scaffolds that encourage perseverance to conquer failure and its resulting negative emotions and starting over with hope, determination, and even enthusiasm.

ACKNOWLEDGMENTS

This research was supported by grants awarded by the U. S. Office of Naval Research (N00014-05-1-0241), the National Science Foundation (ITR 0325428, HCC 0834847), and the Institute of Education Sciences (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 funding agencies.

REFERENCES


APPENDIX: SAMPLE ANALYTICAL REASONING PROBLEM

[SCENARIO]
In a single day, exactly seven people—Anna, Brett, Claudia, Dean, Erica, Frank, & Georgia—come in to interview for a job. Each interview is individual and lasts for one hour. There are two available positions, secretary and messenger. Each person can interview for either position or both. The following conditions apply:
- Anna and Dean do not interview for the same position.
- Frank interviews some time before Claudia.
- No two consecutive interviews are for both positions.
- Three people interview for the secretary position only.
- Georgia interviews fifth.

[SUB QUESTION]
Which of the following could be the order in which the interviews take place?

a) Claudia, Dean, Frank, Brett, Georgia, & Anna
b) Anna, Dean, Frank, Brett, Georgia, & Claudia
c) Brett, Anna, Dean, Frank, Georgia, & Claudia
d) Anna, Frank, Dean, Georgia, Brett, & Claudia
e) Dean, Brett, Frank, Claudia, Anna, & Georgia