Running Head: THE HALF LIFE OF AFFECTIVE STATES

The Half-Life of Cognitive-Affective States during Complex Learning

Sidney D’Mello and Art Graesser

University of Memphis

Corresponding Author:

Sidney D’Mello
209 Dunn Hall
Computer Science Department
University of Memphis
Memphis, TN 38152, USA
Phone: 901-378-0531
Email: sdmello@memphis.edu

Key Words. Cognitive-affective states, temporal dynamics, decay characteristics, learning

Abstract

We investigated the temporal dynamics of students' cognitive-affective states (confusion, frustration, boredom, engagement/flow, delight, and surprise) during deep learning activities. After a learning session with an intelligent tutoring system with conversational dialogue, the cognitive-affective states of the learner were classified by the learner, a peer, and two trained judges at approximately 100 points in the tutorial session. Decay rates for the cognitive-affective states were estimated by fitting exponential curves to time series of affect responses. The results partially confirmed predictions of goal-appraisal theories of emotion by supporting a tripartite classification of the states along a temporal dimension: persistent states (boredom, engagement/flow, and confusion), transitory states (delight and surprise), and an intermediate state (frustration). Patterns of decay rates were generally consistent across affect raters, however, a reverse actor-observer effect was discovered for engagement/flow and frustration. Correlations between decay rates of the cognitive-affective states and several learning measures confirmed existing predictions and yielded some novel findings that have implications for theories of pedagogy that integrate cognitive and affective processes during deep learning.

The Half-Life of Cognitive-Affective States during Complex Learning

Theoretical frameworks that systematically investigate the link between emotions and learning at deeper levels of cognition have been emerging in a number of fields, including psychology (Dweck, 2002; Stein & Levine, 1991), education (Meyer & Turner, 2006), and neuroscience (Immordino-Yang & Damasio, 2007). The fundamental assumption of many of these theories is that affect and cognition are inextricably bound because learning inevitably involves failure and the learner experiences a host of affective responses. Negative emotions (such as confusion, frustration, and sometimes rage) are ordinarily associated with failure, making mistakes, diagnosing what went wrong, struggling with troublesome impasses, and revising plans (see Methods section for definitions of these emotions). Positive emotions (such as delight and excitement) are experienced when tasks are completed, challenges are conquered, insights are unveiled, and major discoveries are made. These threads of negative and positive emotions are consistent with theories that illustrate how emotions are systematically affected by the knowledge and goals of the learner (Csikszentmihalyi, 1990; Dweck, 2002; Mandler, 1999; Meyer & Turner, 2006; Pekrun, Elliot, & Maier, 2006; Stein & Levine, 1991). Consistent with these theories, recent research has identified boredom, engagement/flow, confusion, frustration, delight, and surprise as the major cognitive-affective states that students naturally experience during deep learning and problem solving session that span 30 minutes to 1.5 hours (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Craig, & Graesser, 2009; Graesser & Olde, 2003).

One important aspect of the affect-cognition relationship that has not been adequately addressed is the temporal dynamics of affective experience, called affective chronometry (Davidson, 1998; Rosenberg, 1998). Affective chronometry involves the study of the time course of emotional experiences along temporal dimensions that include *response latency* (i.e., time between the onset and peak intensity of an affective state) and *recovery time* (i.e. time for an emotional experience to dissipate) (Davidson, 1998), which is the focus of this paper.

An understanding of the temporal dynamics of particular classes of affective states is necessary for a satisfactory model that integrates affect with complex learning. For example, confusion is a cognitive-affective state that has been positively correlated with deep thinking and learning (Craig, Graesser, Sullins, & Gholson, 2004). Some learners, such as academic risk takers, like to be challenged with tasks that create short-term confusion, followed by a resolution (Meyer & Turner, 2006). Impasse-driven theories of learning (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003) would suggest that there are learning benefits from episodes of confusion. In these episodes the learner experiences cognitive disequilibrium and is forced to reflect, problem solve and deliberate in an effortful manner in order to restore cognitive equilibrium (Graesser & Olde, 2003; VanLehn et al., 2003). Understanding the temporal dynamics of emotional experiences as they unfold would be necessary to distinguish this type of productive confusion which leads to learning and positive emotions, from being hopelessly confused, which presumably has no pedagogical value.

At this point in science, there is insufficient empirical research to support a categorization of emotions on a temporal dimension. We know that emotions are quite brief (approximately 0.5 – 4 seconds) when they are measured from facial expressions (Ekman, 1984). However, reports of subjective experience of emotion provide much longer estimates ranging from minutes to hours (Frijda, Mesquita, Sonnemans, & Van Goozen, 1991); some of these estimates might be more indicative of moods than emotions per se (Rosenberg, 1998; Watson & Clark, 1994). Additionally, recent evidence from affective neuroscience indicates that there are graded differences in the recovery time from positive and negative affective experiences (Davidson, 1998; Garrett & Maddock, 2001; Hemenover, 2003).

With mixed empirical data and no clear theoretical framework explicitly addressing affective chronometry, we turn to goal-appraisal theories of emotion (Mandler, 1999; Stein & Levine, 1991) for some insights. According to these theories, learners are typically in a *prolonged* state of either (a) engagement/flow as they pursue the superordinate learning goal of mastering the material in the learning environment or (b) disengagement (boredom) when they abandon pursuit of the superordinate learning goal. When they are deeply engaged, they attempt to assimilate new information into existing knowledge schemas. However, when new or discrepant information is detected, attention shifts to the discrepant information, the autonomic nervous systems increases in arousal, and the learner experiences a variety of possible 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 the novelty triggers the achievement of a goal, the emotion is positive, such as delight or contentment.

Surprise that occur in response to novelty and delight when an intermediate learning goal is achieved is expected to be quite brief. In contrast, confusion and frustration occur when the discrepancy or novelty triggers an impasse that blocks the superordinate learning goal and possibly results in the student getting stuck. The learner initiates a subgoal of resolving the impasse through effortful reasoning and problem solving. These states of confusion and frustration address a subgoal, so they should be shorter than the states of engagement/flow and boredom that address the major goal, but are expected to persist longer than the short-lived reactions of delight and surprise. Hence, goal-appraisal theories suggest the following temporal scale in increasing order of persistence: (Delight = Surprise) < (Confusion = Frustration) < (Boredom = Engagement/Flow).

The current study attempted to test these predictions by modeling the temporal dynamics (i.e., decay rates) of naturally occurring cognitive-affective states during a learning session with an Intelligent Tutoring System, AutoTutor. AutoTutor helps learners construct explanations by interacting with them in natural language with adaptive dialogue moves similar to human tutors (Graesser et al., 2004). After the tutorial interaction participants' cognitive-affective states were rated by the learners themselves, untrained peers, and two trained judges. Our primary goal was to answer three questions pertaining to the temporal persistence of the learning-centered cognitive-affective states. First, do the patterns of decay rates align with the predictions of goal-appraisal theories of emotion? Second, are the patterns of decay rates influenced by the affect raters (i.e., self, peers, trained judges). Third, are decay rates related to learning outcomes?

Methods

Participants

The participants were 28 undergraduate students from a mid-south university who participated for extra course credit.

Procedure

Interaction with AutoTutor. Participants completed a multiple-choice pretest and then interacted with AutoTutor for 32 minutes on one of three randomly assigned topics in computer literacy: hardware, Internet, or operating systems. Each of these topics had 12 questions that required about a paragraph of information (3-7 sentences) in an ideal answer. The questions required answers that involved inferences and deep reasoning, such as *why, how, what-if, what if not, how is X similar to Y?* Although each question required 3-7 sentences in an ideal answer, learners rarely give the complete answer in a single conversational turn. Therefore, the tutor scaffolds the construction of an answer by an adaptive dialogue including pumps, hints, prompts, assertions, summaries, and feedback. AutoTutor delivers its dialogue moves via an animated conversational agent that speaks the content of the tutor's turns along with some facial expressions and gestures[1].

During the tutoring session, videos of participants' faces, their computer screens, and their posture patterns (not elaborated here) were recorded. Lastly, after completing the tutoring session, the participants completed a multiple-choice posttest. The pretest and posttest were validated in previous experiments involving AutoTutor (Graesser et al., 2004) and were designed to assess deep levels of knowledge (i.e. causal reasoning, inference, etc.) rather than recall of shallow facts.

Judging Cognitive-Affective States. Similar to a cued-recall procedure (Rosenberg & Ekman, 1994) the judgments for a learner's tutoring session proceeded by playing a video of the face along with the screen capture video of interactions with AutoTutor on a dual-monitor computer system. The screen capture included the tutor's synthesized speech, printed text,

---

[1] A detailed description of AutoTutor is beyond the scope of this brief report. However, please see http:www.autotutor.org for additional information on AutoTutor and https://umdrive.memphis.edu/sdmello/public/AutoTutor-Physics-FullSubtopic/AutoTutor-Physics-FullSubtopic.htm for a demo of AutoTutor.

students' responses, dialogue history, and images, thereby providing the context of the tutorial interaction.

Raters were instructed to make judgments on what affective states were present at any moment during the tutoring session by manually pausing the videos. They were also instructed to make judgments at each 20-second interval where the video automatically stopped. If the learner was experiencing more than one affective state, raters were instructed to mark each state and indicate which was most pronounced. However, only the first choice (more prominent) affective states were included in the subsequent analyses.

Raters were provided with a checklist of seven states for them to mark along with definitions of the states. Boredom was defined as being weary or restless through lack of interest. Confusion was defined as a noticeable lack of understanding, whereas engagement/flow was a state of interest that results from involvement in an activity. Frustration was defined as dissatisfaction or annoyance. Delight was a high degree of satisfaction. Surprise was wonder or amazement, especially from the unexpected. Neutral was defined as no apparent emotion or feeling. Hence, judgements were made on the basis of the learners' facial expressions, contextual cues via the screen capture, and the definitions of the cognitive-affective states.

Four sets of judgments were made for the observed affective states of each AutoTutor session. First, for the *self* judgments, the learner watched his or her own session with the tutor immediately after having interacted with AutoTutor. Second, for the *peer* judgments, each learner came back a week later to watch and judge another learner's session. Finally, there were two trained judges: undergraduate research assistants who were trained extensively on AutoTutor's dialogue characteristics (i.e., the context) and how to detect facial action units according to Ekman’s Facial Action Coding System (Ekman & Friesen, 1978). The two trained judges judged all sessions separately.

Analysis of Agreement Across Judges

The affect judgment procedure yielded 2,967 self judgments, 3,012 peer judgments, and 2,995 and 3,093 judgments for the two trained judges. We computed proportional agreement scores for the six pairs of raters by independently considering pairs of raters and dividing their number of agreements by the total number of observations. Proportional agreement scores for the six rater pairs were: self-peer (.279), self-judge1 (.364), self-judge2 (.330), peer-judge1 (.394), peer-judge2 (.368), and judge1-judge2 (.520). These scores indicate that the trained judges had the highest agreement, the self-peer pair had the lowest agreement, and the other pairs of judges were in between. Another finding is that there are actor-observer differences in the agreement scores. The average actor-observer agreement was .324 (i.e., average of self-peer, self-judge1, and self-judge2), which is lower than the average observer-observer agreement score of .427 (i.e., average of peer-judge1, peer-judge2, judge1-judge2).

The overall low agreement scores highlight the difficulty of judging the cognitive-affective states that spontaneously emerge during learning. Our use of multiple raters is justified by the fact that there is no clear gold standard to declare what the learner’s cognitive-affective states truly are (Graesser et al., 2006). Is it the self, the untrained peer, the trained judges, or physiological instrumentation? A neutral, but defensible position is to independently consider ratings of the different judges, thereby allowing us to examine patterns that generalize across judges as well as patterns that are sensitive to individual judges. This strategy was adopted in the current paper.

Results and Discussion

We investigated the manner in which the affective states persisted over time by estimating decay rates for each of the states. Our analyses proceeded by considering several time intervals (from 0 seconds to 60 seconds, with 20 second increments) and computing the probability that each state persisted for each time interval, i.e., $Pr(E_t = E_{t+20})$. For example, consider the following sample time series of self-reported states for a given participant, $\underline{B^{20}\ B^{40}\ B^{60}}$ X^{80} X^{100} X^{120} X^{140} X^{160} $\underline{B^{180}\ B^{200}}$ X^{220} $\underline{B^{240}}$ X^{260}; B refers to boredom, X some other state (confusion, delight, etc.,), and the superscripts $(20, 40$, etc) refer to timestamps in seconds. Let us focus on the *unique* episodes of boredom in this sample time series. An episode of boredom $(E_t = B)$ is unique if the previous state is not boredom $(E_{t-20} \neq B)$. In line with this definition, there are three unique episodes of boredom in this time series (B^{20}, B^{180}, and B^{240})**.** After identifying the three boredom episodes, we compute the probability that an episode of boredom persists for at least 0, 20, 40, and 60 seconds[2]. Persistence at zero seconds is the simplest case because $Pr(E_{t=0} = B) = 1$. Next we compute the probability that boredom persists at the next time step, which is 20-second later. Here, $Pr(E_{t\geq 20} = B) = \frac{2}{3} = .67$, because only the first and the second instances of boredom (B^{20}, B^{180}) persisted for at least 20 seconds. Similarly, $Pr(E_{t\geq 40} = B) = \frac{1}{3} = .33$ and $Pr(E_{t\geq 60} = B) = \frac{0}{3} = 0$. This process yielded the time series depicted in Figure 1A.

---

[2] 20 seconds was chosen as the upper boundary because an examination of the data indicated that affective states rarely persisted for more than 20 seconds.

In this fashion separate time series were created for each affective state as reported by each rater for each participant. Although one would expect 672 time series (28 participants × 6 affective states × 4 raters), we obtained 574 time series because all participants did not experience all states.

A visual inspection of the data revealed that an exponential model would appropriately capture the decay characteristics of the states. According to an exponential decay model, the probability that a state will persist (i.e. not transition into neutral or another state) at time *t* is $y(t) = b_o e^{-b_1 t}$, where b_o is the initial value (i.e. the probability that the state existed at time t = 0) and b_1 is the decay rate. According to Figure 1A, $b_o = 1$ for all states, because the probability that a state will persist at 0 seconds is 1.0. Our analyses proceeded by fitting exponential curves to estimate the decay rate (b_1) of each state after fixing the initial value b_o at 1.

Separate models were fit to each of the 574 time series and the decay rates were retained for subsequent analyses. An analysis of the quality of the fit indicated that the exponential models provided a good fit to the data because the grand mean R^2 was .873. Mean R^2 values for the individual states were: R^2_{boredom} = .837, $R^2_{\text{confusion}}$ = .891, R^2_{delight} = .936, R^2_{flow} = .844, $R^2_{\text{frustration}}$ = .896, and R^2_{surprise} = .897 (these means were obtained by averaging across participants and affect raters).

INSERT FIGURE 1 ABOUT HERE

Patterns of Decay Rates Across Affective States and Affect Raters

The decay rates were analyzed with a linear mixed-effects full-factorial repeated-measures model (Pinheiro & Bates, 2000), with affect and rater as fixed factors and participants as a random factor. The model yielded a significant main effect for affect, $F(1, 5) = 46.4$, p <

.001, rater $F(1, 3) = 12.1$, $p < .001$, and a significant affect × rater interaction, $F(1, 15) = 4.14$, $p < .001$.

When compared to the null model that suggests that decay rates are similar for all affective states, the significant main effect for affect indicates that the decay rate of at least one affective state is different from the others. Instead of performing posthoc tests and reporting multiple *p*-values, we identify differences in decay rates associated with this main effect from the means and 95% confidence intervals presented in Table 1. An examination of Table 1 revealed the following pattern in the data: (Delight = Surprise) < Frustration < (Confusion = Engagement/Flow = Boredom). Lower negative numbers for b_1 are indicative of more rapid decay, hence, these results indicate that delight and surprise decay the fastest, confusion, flow, and boredom the slowest, and the decay rate of frustration lies between these two extremes (see Figure 1B).

The pattern of means for the main effect of rater (see Table 1) is consistent with general consensus among raters, but with one exception. Decay rates associated with self-reported affective states were similar to decay rates of the peers and the first trained judge. However, the decay rates associated with rating by judge 2 are more rapid than decay rates obtained from ratings of the other judges.

INSERT TABLE 1 ABOUT HERE

The significant, but comparatively less robust, affect × rater interaction presented in Table 2 indicates that decay rates for particular affective states are influenced by the affect raters. It appears that decay rates for boredom and confusion were similar across raters with the exception that the second trained judges' decay rates were comparatively less variable and higher for boredom while the peers' decay rates were more variable and lower for confusion. Decay rates for delight and surprise were also similar for ratings provided by the self, peers, and trained judge 1, although there is considerably more variability for these states. Judge 2 deviated from this pattern because this judges' decay rates were more rapid for delight and surprise. In general, with the exception of one rater (peer for confusion, judge 2 for other three states), decay rates associated with the actors (i.e., self judgments) and observers (i.e., peers and trained judges) were similar for boredom, confusion, delight, and surprise, at least within the range of the confidence intervals (see Figure 2).

More interesting patterns pertaining to the affect × rater interaction emerged for engagement/flow and frustration. The actor-observer effect is particularly prevalent for these states (see Table 2 and Figure 2). Decay rates associated with self-reported engagement/flow were less variable and lower than decay rates for the peers and both trained judges, which were on par with each other. A similar pattern was observed for frustration, with the exception that judge 2's decay rates were more rapid than the peers and judge 1. So the pattern for frustration is: Self < [Peer = Judge 1] < Judge 2.

INSERT TABLE 2 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

Correlations between Decay Rates and Learning Measures

We performed a set of correlational analyses between the decay rates of the six cognitive-affective states (averaged across the four judges) and four performance measures including prior knowledge, learning efficiency, acquired knowledge, and transferred knowledge. These measures were computed from the multiple-choice tests that were administered before and after the tutorial session. The tests were designed to assess deep levels of knowledge with questions that required inferences and deep reasoning (e.g., *why, how, what-if, what if not, how is X similar to Y?*), instead of simply asking students to recall previously presented information, definitions, and facts (i.e., shallow measures).

Prior knowledge is the proportion of questions answered correctly on the pretest. *Learning efficiency* is the proportion of problems the learners completed in the 32 minute tutorial session. *Acquired knowledge* is proportional learning gains on topics for which the students received tutoring (i.e., one out of the three computer literacy topics), while *transferred knowledge*, an important measure of deep comprehension, is proportional learning gains on related topics for which the student received *no* tutoring (i.e., remaining two computer literacy topics). Proportional learning gains measure the extent of improvement over prior knowledge and are computed as: (posttest scores – pretest scores) / (1 - pretest scores).

A 6 × 4 (affect × performance measure) correlational matrix is presented in Table 3. Although we tested the significance of the correlational coefficients, our small sample size of 28 participants does not yield sufficient statistical power to detect small ($r \approx .1$) and medium sized effects ($r \approx ..3$). Hence, in addition to discussing significant effects we also consider non-significant correlations of .3 or higher to be meaningful because these are likely to be significant with a larger sample.

INSERT TABLE 3 ABOUT HERE

The analyses yielded a number of interesting patterns between the decay rates and the performance measures. It appears that prior knowledge was negatively correlated with the decay rate of engagement/flow, but not with any of the other states. Hence, the more knowledgeable learners are less likely to persist in a state of heightened engagement. This finding is plausible because an appropriate balance between challenge and skill is essential to maintaining the zone of flow (Csikszentmihalyi, 1990). The knowledgeable learners are apparently not being sufficiently challenged, which is what could be expected from a tutoring system designed to help learners with little to no prior knowledge.

Another finding was that acquired knowledge was positively correlated with engagement/flow, but negatively correlated with boredom, frustration, and confusion. Confusion was also the only state that was negatively correlated with learning efficiency. One interpretation of this pattern is that confused learners acquire less knowledge, presumably because they cover less content due to the time required to work thorough their confusion. On the other hand, learners that persist in boredom and frustration cover the same amount of content but with no mastery. Increasing time on task for confused learners might produce learning gains comparable to the engaged learners, however, this hypothesis requires empirical confirmation.

Perhaps the most important finding was that transferred knowledge was positively correlated with confusion, but not any of the other states. Persisting in confusion helps with knowledge transfer because confusion is a state that requires learners to stop and think and problem solve. This finding implies that confusion is one precursor to deep learning (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005), and is consistent with theories that highlight the merits of impasses during learning (VanLehn et al., 2003), desirable difficulties (Bjork & Linn, 2006), and with models that help students learn how to overcome failure from getting stuck (Burleson & Picard, 2007).

General Discussion

We investigated the temporal dynamics of cognitive-affective states during complex learning. Our results offer a finer grained understanding of affect-learning relationships than previous theories that attempted to distinguish emotions from moods and affective traits

(Rosenberg, 1998; Watson & Clark, 1994), or positive and negative emotions (Davidson, 1998; Garrett & Maddock, 2001), or views that allege that emotions last for only a few seconds (Ekman, 1984), or minutes or hours (Frijda et al., 1991). The subsequent discussion focuses on thee of our major findings followed by some of the pedagogical implications of our results.

One fundamental conclusion from this research is that there are differences in the patterns of decay rates of the cognitive-affective states. In particular, our results indicated that delight and surprise are transitory states that dissipate almost immediately, boredom, engagement/flow, and confusion are persistent states that tend to linger for a while, and frustration is an intermediate state. These results seem to partially confirm predictions stemming from goal-appraisal theories of emotion (Mandler, 1999; Stein & Levine, 1991). These theoretical frameworks would suggest that boredom and engagement/flow are persistent states that are aligned with the superordinate goal of mastering the material, confusion and frustration are shorter because they are aligned with the subgoal of handling impasses, and delight and surprise are transitory states. These predictions were generally supported with the exception that confusion was categorized as a persistent state rather than a state of intermediate duration.

Another important finding pertains to the discovery of an affect × rater interaction for the pattern of decay rates. While decay rates derived from affect ratings provided by the four raters were generally consistent for boredom, confusion, delight, and surprise, actor-observer differences were particularly prevalent for engagement/flow and frustration. Theories that emphasis actor-observer differences would predict that self reported states would be more transient because actors (learners) presumably attribute their emotions to malleable situational factors. On the other hand, observers (peers and trained judges) might make attributions to stable dispositional factors of the learner, thereby considering the affective states to be more enduring

(Jones & Nisbett, 1971). In contrast to this view, our results support a reverse actor-observer effect for engagement/flow and frustration. These states were considered to be persistent when the affect ratings were provided by the learners themselves (i.e., the actors), however, the observers (peers and trained judges) viewed these states as being more transient.

One explanation for this pattern might lie in the expression of engagement/flow and frustration on the face. The state of engagement/flow, does not appear to have overly diagnostic facial features (McDaniel et al., 2007), hence observers might consider engagement/flow to have dissipated when they cannot detect it on the face. Frustration does have discriminating facial correlates, however, there is some evidence that learners attempt to disguise displays of frustration (Craig, D'Mello, Witherspoon, & Graesser, 2008; McDaniel et al., 2007), thereby making it difficult for observers to detect this emotion from the face. Hence, the absence of persistent facial cues for these states might explain why the observers considered engagement/flow and frustration to be more ephemeral. This problem of restricted facial cues has less of an impact on self-reported ratings because in addition to videos of their faces, learners can also consult recent memories of the tutorial session when making their affect judgments.

Our third important finding was that the decay rates of boredom, engagement/flow, confusion, and frustration were correlated with our learning measures. As predicted by theories that emphasize the importance of affect to deep learning (Dweck, 2002; Immordino-Yang & Damasio, 2007; Meyer & Turner, 2006; Stein & Levine, 1991), persisting in a flow state is critical for learning, while wallowing in boredom and frustration is detrimental to learning. One somewhat controversial discovery was the confusion was a powerful predictor of knowledge transfer, which is presumably the most important measure of deep learning. When coupled with the categorization of confusion as a persistent state, this finding highlights the importance of confusion to deep learning activities. It appears that upon activation, confusion adopts a persistent temporal quality while learners are in a state of cognitive disequilibrium. Cognitive equilibrium is normally restored after thought, reflection, problem solving, and other effortful cognitive activities. It is not the confusion itself, but the effortful cognitive activities aimed at resolving the confusion that is beneficial to learning.

Our results have important implications for theories of pedagogy. Clearly, due to our small sample size, correlational methodology, and relatively coarse affect measurement, more research is needed to substantiate some of the trends we have found in our investigations of affect-learning relationships. However, available data support a number of claims. Although persistent episodes of confusion appear to be beneficial to learning, wallowing in the negative states of boredom and frustration has detrimental effects. Students who remained bored and frustrated were not effective learners. In contrast, deep learning is much higher in conditions that present challenges to inspire deep inquiry. Therefore, a promising strategy to promote opportunities for deep learning is to jolt students out of their perennial state of blasé comprehension by presenting challenges with contradictions, incongruities, anomalies, system breakdowns, and difficult decisions. Pedagogical strategies that capitalize on the natural steps of making mistakes (feeling confused), diagnosing what went wrong (without becoming frustrated), discovering a relevant insight (with delight), and starting over again (with hope, determination, and maybe even enthusiasm) are expected to accompany model-based reasoning strategies that are essential for learners of complex subject matter.

References

Baker, R., D'Mello, S., Rodrigo, M., & Graesser, A. (2010). Better to be frustrated than bored: The incidence and persistence of affect during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68 (4)*, 223-241.

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.

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.

Craig, S., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive-affective states during learning. *Cognition & Emotion, 22*(5), 777-788.

Craig, S., Graesser, A., Sullins, J., & Gholson, J. (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: Harper and Row.

D'Mello, S., Craig, S., & Graesser, A. (2009). Multi-method assessment of affective experience and expression during deep learning. *International Journal of Learning Technology, 4*(3-4), 165-187.

Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition & Emotion, 12*, 307-330.

Dweck, C. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.

Ekman, P. (1984). Expression and the nature of emotion. In K. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319-344). Hillsdale, NJ: Erlbaum.

Ekman, P., & Friesen, W. (1978). *The Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena, or emotions, sentiments, and passions. In K. Strongman (Ed.), *International review of emotion and motivation* (pp. 187-225). New York: Wiley.

Garrett, A. S., & Maddock, R. J. (2001). Time course of the subjective emotional response to aversive pictures: Relevance to fMRI studies. *Psychiatry Research, 108*(1), 39-48.

Graesser, A., Lu, S., Olde, B., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition, 33*, 1235-1247.

Graesser, A., 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., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. (2006). *Detection of emotions during learning with AutoTutor.* Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada.

Graesser, A., & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95*(3), 524-536.

Hemenover, S. H. (2003). Individual differences in rate of affect change: Studies in affective chronometry. *Journal of Personality and Social Psychology, 85*, 121-131.

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.

Jones, E., & Nisbett, R. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. New York: General Learning Press.

Mandler, G. (1999). Emotion. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive science. Handbook of perception and cognition* (2nd ed.). San Diego, CA: Academic Press.

McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., & Graesser, A. (2007). Facial Features for Affective State Detection in Learning Environments. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 467-472). Austin, TX: Cognitive Science Society.

Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377-390.

Pekrun, R., Elliot, A., & Maier, M. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*(3), 583-597.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.

Rosenberg, E. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*(3), 247-270.

Rosenberg, E., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion, 8*(3), 201-229.

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

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209-249.

Watson, D., & Clark, L. A. (1994). Emotions, moods, traits, and temperaments: Conceptual distinctions and empirical findings. In P. Ekman & J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 89-93). New York: Oxford University Press.

Authors Notes

This research was supported by the National Science Foundation (REC 0106965, ITR 0325428, HCC 0834847). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

Table 1.

Descriptive Statistics and Confidence Intervals for Main Effects of Affect and Rater

Factor	**Mean**	**Stdev**	**95% CI** *Lower*	*Upper*
Affect				
Boredom	-.054	.058	-.076	-.031
Confusion	-.038	.028	-.049	-.027
Delight	-.292	.218	-.383	-.200
Flow	-.047	.058	-.070	-.025
Frustration	-.156	.126	-.205	-.108
Surprise	-.303	.194	-.380	-.226
Rater				
Self	-.075	.061	-.099	-.051
Peer	-.103	.104	-.143	-.062
Judge 1	-.108	.070	-.135	-.081
Judge 2	-.163	.081	-.194	-.132

Table 2.

Descriptive Statistics and Confidence Intervals for Affect × Rater Interaction

Affect	**Rater**	**Mean**	**Stdev**	**95% CI**	
				Lower	*Upper*
Boredom	Self	-.072	.160	-.137	-.007
	Peer	-.047	.112	-.090	-.003
	Judge 1	-.070	.158	-.132	-.008
	Judge 2	-.030	.008	-.033	-.027
Confusion	Self	-.035	.016	-.041	-.029
	Peer	-.057	.111	-.100	-.014
	Judge 1	-.029	.012	-.034	-.024
	Judge 2	-.030	.008	-.033	-.027
Delight	Self	-.168	.247	-.291	-.045
	Peer	-.283	.288	-.431	-.135
	Judge 1	-.260	.266	-.375	-.144
	Judge 2	-.422	.263	-.538	-.305
Flow	Self	-.027	.010	-.031	-.023
	Peer	-.050	.119	-.099	-.001
	Judge 1	-.048	.112	-.092	-.005
	Judge 2	-.054	.116	-.101	-.007
Frustration	Self	-.035	.015	-.041	-.029
	Peer	-.146	.217	-.237	-.054
	Judge 1	-.113	.195	-.195	-.030
	Judge 2	-.365	.282	-.490	-.239
Surprise	Self	-.199	.262	-.315	-.083
	Peer	-.270	.290	-.406	-.135
	Judge 1	-.348	.294	-.499	-.197
	Judge 2	-.478	.247	-.610	-.347

Table 3.

Correlations between Decay Rates and Learning Measures

Affect	**Prior Knowledge**	**Learning Efficiency**	**Acquired Knowledge**	**Transferred Knowledge**
Boredom	.171	-.014	.-.299	.256
Confusion	-.236	*-.362	**-.463	***.601
Delight	-.002	-.077	-.005	-.240
Flow	**-.424	.206	.313	-.125
Frustration	.239	-.005	-.316	.154
Surprise	-.134	-.012	-.041	-.007

Notes. . * p < .1 ** p < .05, *** p < .01.

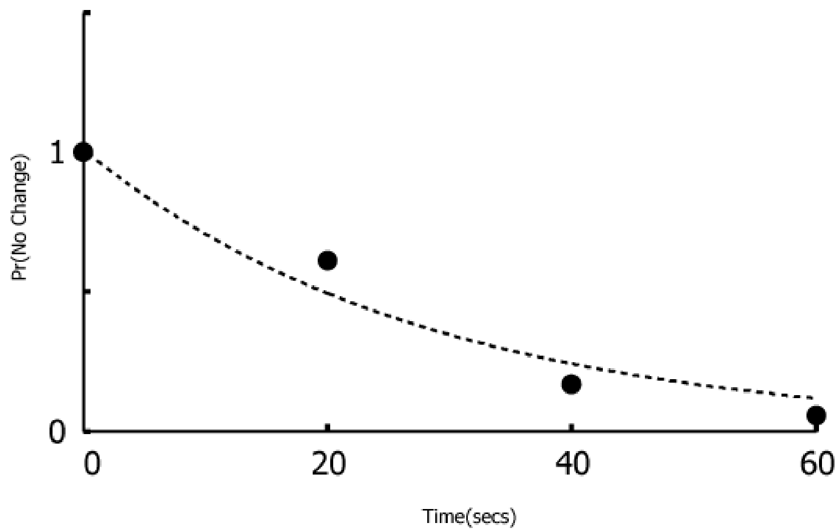
Figure Captions

Figure 1 (A). Sample time series illustrating exponential curve fit (B). Exponential decay curves for the main effect of affect

Figure 2. Exponential decay curves for the affect × rater interaction

Figure 1.

A. Sample time series with exponential decay curve fit



B. Exponential Decay Models

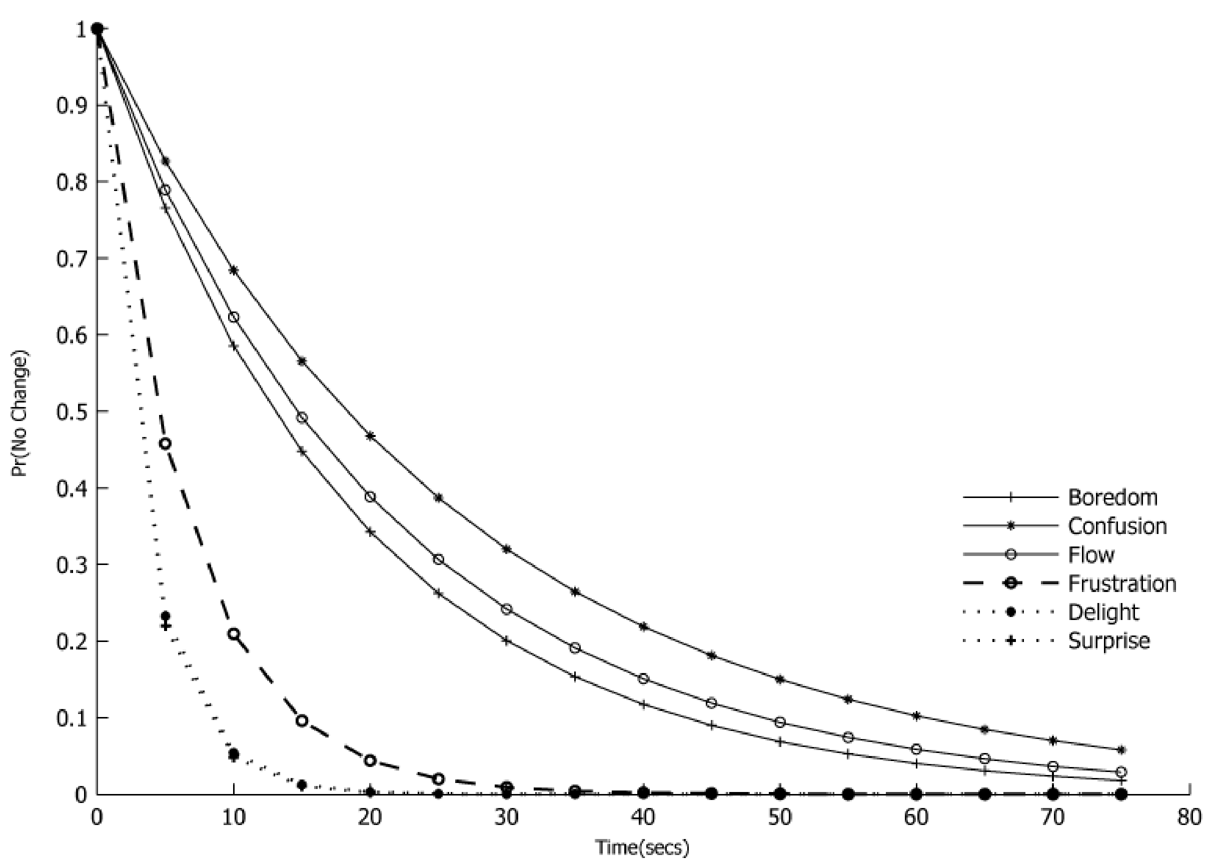
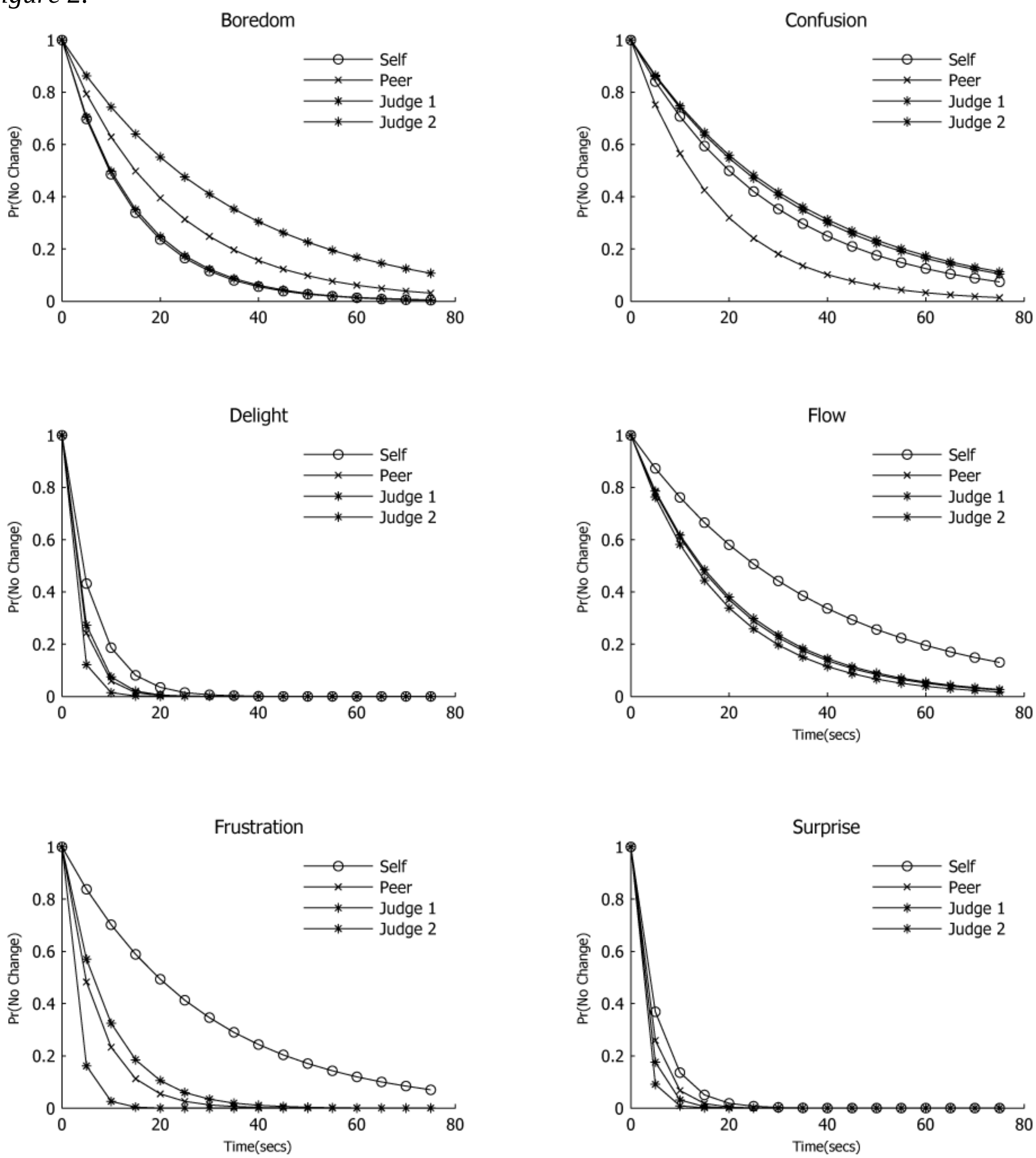


Figure 2.


Boredom
Confusion
Delight
Flow
Frustration
Surprise
Self
Peer
Judge 1
Judge 2
Pr(No Change)
Time(secs)