Inducing and Tracking Confusion with Contradictions during Critical Thinking and Scientific Reasoning

Blair Lehman[1], Sidney D'Mello[1], Amber Strain[1], Melissa Gross[1], Allyson Dobbins[1], Patricia Wallace[2], Keith Millis[2], and Art Graesser[1]

[1] Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152
[balehman, sdmello, dchuncey, magross, bdbbins1, graesser]@memphis.edu,
[2] Department of Psychology, Northern Illinois University, Dekalb, Illinois 60115
[pwallace, kmillis]@niu.edu

Abstract. Cognitive disequilibrium and its affiliated affective state of confusion have been found to be beneficial to learning due to the effortful cognitive activities that accompany their experience. Although confusion naturally occurs during learning, it can be induced and scaffolded to increase learning opportunities. We addressed the possibility of induction in a study where learners engaged in trialogues on critical thinking and scientific reasoning topics with animated tutor and student agents. Confusion was induced by staging disagreements and contradictions between the animated agents, and the (human) learners were invited to provide their opinions. Self-reports of confusion and learner responses to embedded forced-choice questions indicated that the contradictions were successful at inducing confusion in the minds of the learners. The contradictions also resulted in enhanced learning gains under certain conditions.

Keywords: Confusion, cognitive disequilibrium, contradiction, affect, tutoring, intelligent tutoring systems, learning.

1 Introduction

Connections between complex learning and emotions have received increasing attention in the fields of psychology [1-3], education [4-6], neuroscience [7], and computer science [8-11]. An understanding of affect-learning connections is needed to design engaging educational artifacts that range from affect-sensitive intelligent tutoring systems (ITSs) on technical material to entertaining media [12, 13].

The fundamental assumption behind much of this research is that affect and cognition are inextricably bound and fundamental to learning. This assumption is reasonable if one realizes that learning inevitably involves failure and a host of affective responses. Negative emotions (e.g., confusion, irritation, frustration, anger, and sometimes rage) are ordinarily associated with making mistakes, diagnosing what went wrong, and struggling with impasses. Positive emotions (e.g., engagement, flow, delight, excitement, and eureka) are experienced when tasks are completed, challenges are conquered, and major discoveries are made.

Importantly, the relationship between affect and learning is more complex than a simple model which posits that positive emotions facilitate learning while negative emotions hinder learning. Perhaps one of the most significant and counterintuitive findings pertains to the role of confusion in promoting deep learning. Confusion occurs when students get stuck; are confronted with a contradiction, anomaly, or system breakdown; and are uncertain about what to do next. Confusion provides an opportunity for learning because it triggers active problem solving and reasoning, a view that is consistent with impasse-driven theories of learning [14-16].

Evidence for impasse-driven learning can be found in early work on skill acquisition and learning [14-16]. For example, in an analysis of over 100 hours of human-human tutorial dialogues, VanLehn et al. [16] reported that comprehension of physics concepts was rare when students did not reach an impasse, irrespective of the quality of explanations provided by the tutor. There is also some evidence that confusion is positively correlated with learning due to the activities associated with its resolution (i.e., effortful elaboration and causal reasoning during problem solving) [17, 18]. These activities involve desirable difficulties [19], which inspire greater depth of processing, more durable memory representations, and more successful retrieval [20].

In our view, the complex interplay between events that trigger confusion coupled with effortful impasse-resolution processes is the key to promoting deep learning. Learning is presumably not directly caused by confusion, but rather by the cognitive activities that accompany its experience. The benefits of impasses and confusion can only be leveraged in a learning environment (LE) if three conditions are met: (1) the LE has events that *induce* confusion; (2) the LE can detect and *track* the associated confusion; and (3) the LE *regulates* confusion in a way that maximizes learning. The focus of this paper is to systematically explore methods to induce confusion in the learner so this paper will mainly focus on research activities to advance this goal.

We describe a study in which confusion was experimentally induced in an LE with two pedagogical agents that engaged in a trialogue with the human learner. The two agents served as the medium through which confusion is induced over the course of learning critical thinking and scientific reasoning skills such as designing and evaluating research studies. We focused on two research questions. First, can confusion be induced when the agents contradict each other and ask the human learner to intervene? More specifically, will confusion be induced if one agent presents accurate information and the other presents inaccurate information? Second, what are the indicators of the induced confusion?

2 Method

2.1 Manipulation

We experimentally induced confusion with a contradictory information manipulation over the course of learning concepts in critical thinking (e.g., random assignment, experimenter bias). This is achieved by having the tutor and student agents stage a disagreement on an idea and eventually invite the human to intervene (note that student agent refers to an animated agent, the actual learner is referred to as participant or learner). The contradiction is expected to trigger conflict and force the participant to reflect, deliberate, and decide which opinion has more scientific merit.

Contradictions were introduced during trialogues identifying flaws in sample research studies. Some studies had subtle flaws while others were flawless. There were four contradictory information conditions. In the *true-true* condition, the tutor agent presented a correct opinion and the student agent agreed with the tutor; this is the no contradiction control. In the *true-false* condition, the tutor presented a correct opinion and the student agent disagreed by presenting an incorrect opinion. In contrast, it was the student agent who provided the correct opinion and the tutor agent who disagreed with an incorrect opinion in the *false-true* condition. Finally in the *false-false* condition, the tutor agent provided an incorrect opinion and the student agent agreed. It should be noted that all misleading information was corrected over the course of the trialogues and participants were fully debriefed at the end of the experiment.

The excerpt in Table 1 is an example trialogue between the two agents and the human learner. This is an excerpt from the *true-false* condition, where the tutor agent (Dr. Williams) and the student agent (Chris) are discussing a flawed study with Bob (the human learner).

Table 1. Excerpt of trialogue from *true-false* condition

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

2.2 Participants and Design

Participants were 32 undergraduate students from a mid south university in the US and participated for course credit. Data from one participant was discarded due to experimenter error. The experiment had a within-subjects design with four conditions (*true-true, true-false, false-true, false-false*). Participants completed two learning sessions in each of the four conditions with a different critical thinking topic in each session (8 in all). Order of conditions and topics and assignment of topics to conditions was counterbalanced across participants with a Graeco-Latin Square.

2.3 Procedure

The experiment occurred over two phases: (1) knowledge assessments and learning sessions and (2) a retrospective affect judgment protocol.

Knowledge Tests. Critical thinking knowledge was tested before and after learning sessions (pretest and posttest, respectively). Each test had 24 multiple-choice questions, three questions per concept (control group, construct validity, correlational studies, experimenter bias, generalizability, measure quality, random assignment, replication). There were three types of test items: definition, function, and example. Random assignment, for example, was assessed with the following questions: "Random assignment refers to __" (definition), "Random assignment is important because __" (function), and "Which study most likely did not use random assignment." (example). There were two alternate test versions and assignment was counterbalanced across participants for pretest and posttest.

Learning Sessions. First, participants signed an informed consent and then completed the pretest. Next, participants read a short introduction to critical thinking topics to familiarize them with the terms that would be discussed. Participants then began the first of eight learning sessions. A webcam and a commercially available screen capture program (Camtasia Studio^{TM}) recorded participants' face and screen, respectively, during the learning sessions.

Each learning session began with a description of a sample research study. Participants read the study and then began a trialogue with the agents. The discussion of each study involved four trials. For example, in Table 1 dialogue turns five through eight represent one trial. Each trial consisted of the student (turn 5) and tutor (turn 6) agents asserting opinions, prompting participants to intervene (turn 7), and obtaining participants' responses (turn 8).

This cycle was repeated in each trial, with each trial becoming increasingly more specific about the scientific merits of the study. The trialogue in Table 1 discusses a study that does not properly use random assignment. Trial 1 broadly asks if students would change their behavior based on the results of the study (turns 1-3), while Trial 2 addressed whether or not a problem is present (turns 5-8). Trial 3 began to specifically address the problematic part of the study, "Do the experimenters know that the two groups were equivalent?". Finally, Trial 4 directly addressed the use of random assignment, "Should the experimenters have used random assignment here?". Participants then completed the posttest after discussing the eight studies.

Retrospective Affect Judgment Protocol. Participants then completed a retrospective affect judgment protocol [21]. Videos of participants' face and screen were synchronized and participants made affect ratings while viewing these videos. Participants were provided with a list of affective states (anxiety, boredom, confusion, curiosity, delight, engagement/flow, frustration, surprise, and neutral) with definitions. Affect judgments occurred at 13 pre-specified points (e.g., after contradiction presentation, after forced-choice question, after learner response) in each learning session (104 in all). In addition to these pre-specified points, participants were able to manually pause the videos and provide affect judgments at any time.

3 Results and Discussion

We hypothesized that contradictory information would induce confusion in learners. To investigate this hypothesis, the experimental conditions (*true-false*, *false-true*, and *false-false*) were compared to the no-contradiction control condition (*true-true*) in two analyses: (1) self-reported levels of confusion and (2) responses to forced-choice questions. In addition, learning gains in experimental conditions were compared to the control condition.

3.1 Retrospective Self-report Confusion Ratings

Although a total of eight affective states were tracked, the present analysis only focuses on confusion because this is the primary dependent measure of interest. The analyses proceeded by computing proportional scores for self-reported confusion ratings in each condition. Paired sample *t*-tests indicated that there was significantly more confusion in the *true-false* condition (M = .06, SD = .10) than the *true-true* condition (M = .04, SD = .06), $t(30) = 2.02$, $p = .03$. However, the other experimental conditions (*false-true* and *false-false*) were not associated with significantly higher levels of confusion than the control (M = .04, SD = .06 and M = .05, SD = .08, respectively). These findings suggest that contradiction between agents can induce some confusion in learners. The success of contradiction, however, does appear to be tempered by who (tutor vs. student) takes the correct vs. incorrect position.

3.2 Tracking Uncertainty via Performance on Forced-Choice Questions

Self-reports are one viable method to track confusion. However, this measure is limited by the learner's sensitivity and willingness to report their confusion levels. A more subtle and promising measure of confusion and uncertainty is to assess learner responses to forced-choice questions following contradictions by the animated agents (see turns 3 and 8 in Table 1). Since these questions adopted a two-alternative multiple-choice format, random guessing would yield a score of 0.5. One-sample *t*-tests comparing learner responses to a chance value of 0.5 revealed the following pattern of performance: (a) *true-true* (M = .76, SD = .19) and *true-false* (M = .60, SD = .19) conditions were significantly greater than chance, (b) *false-true* (M = .45, SD = .26) was statistically indistinguishable from chance, and (c) *false-false* (M = .35, SD = .31) was significantly lower than chance. An ANOVA revealed the following pattern of response correctness across conditions: *true-true* > *true-false* > *false-true* > *false-false*, $F(3,90) = 16.9$, $Mse = .059$, $p < .001$, partial-eta squared = .39.

These results suggest that contradictions successfully evoked uncertainty. The magnitude of uncertainty was dependent upon the source and severity of the contradiction. Uncertainty is low when both agents are correct and there is no contradiction (*true-true*), but increases when one agent is incorrect. Uncertainty is greater when the tutor is incorrect (*false-true*) compared to when the tutor is correct (*true-false*), presumably because this challenges conventional norms. Finally, uncertainty is greatest when both agents are incorrect, even without a contradiction (*false-false*). Hence, uncertainty is maximized when learners detect a clash between their knowledge and the agents' responses. This uncertainty is a likely opportunity to scaffold deep comprehension by forcing learners to stop and think.

3.4 Learning Gains

Paired sample one-tail *t*-tests comparing the proportional learning gains in experimental conditions to the control condition were separately conducted for each question type (i.e., definition, function, example). Pretest and posttest scores were computed as the proportion of questions answered correctly. Proportional learning gains were computed as (posttest – pretest)/(1-pretest).

The results indicated that contradictions differentially impacted shallow and deep learning gains. For definition questions, the most shallow level, learning gains were marginally higher in the *true-true* condition (M = .24, SD = .59) than the *false-true* condition (M = .12, SD = .44), $t(30)$ = 1.87, p = .08, d = .22. However, this pattern was reversed for example questions that assess understanding at deeper levels. The *false-true* condition was marginally higher (M = .24, SD = .60) than the *true-true* condition (M = .00, SD = .64), $t(30)$ = 1.84, p = .08, d = .39. There were no significant learning gain differences for functional questions and with the other experimental conditions (*true-false*, *false-false*).

4 General Discussion

While recent research has identified a set of affective states that are very relevant to learning (e.g., boredom, engagement/flow, confusion, frustration, anxiety, curiosity), the question still remains of how to coordinate affective and cognitive processes to increase learning gains. The strategy we have adopted involves inducing particular affective states and subsequently helping learners regulate these affective states over the course of the session. The present paper reported on one such effort, specifically, on confusion induction during learning. Through the presentation of contradictory information, we were able to successfully induce confusion in learners. Both self-reports of confusion and learner responses to forced-choice questions showed that conditions with a contradiction induced more confusion than the no-contradiction control condition. Learner responses, however, may serve as a more effective and unbiased method to track confusion and uncertainty because learners might be hesitant to report that they are confused or might not be consciously aware of their confusion.

We did not expect impressive learning gains because confusion was only induced and not appropriately scaffolded in this preliminary study. Nevertheless, there were modest improvements in learning deeper content (example questions) in the *false-true* condition. This *false-true* condition was associated with chance-level responses to prompts (*intermediate confusion*), while responses were above chance for the *true-false* condition (*insufficient confusion*) and below chance for the *false-false condition* (*hopeless confusion*). Hence, the *false-true* condition which is associated with just the right level of confusion appears to be the most promising avenue for future research.

Since we have had some success in inducing confusion and uncertainty, the next step is to implement interventions that will make use of these learning opportunities. A learning environment (LE) that detects learner confusion has a variety of paths to pursue. The LE might want to keep the learner confused (i.e. in a state of cognitive disequilibrium) and leave it to the learner to actively deliberate and reflect on how to restore equilibrium. This view is consistent with a Piagetian theory [22] that stipulates that students need to experience cognitive disequilibrium for a sufficient amount of time before they adequately deliberate and reflect via self-regulation. If so, the LE should give indirect hints and generic pumps to get the student to do the talking when floundering. Alternatively, Vygotskian theory [23] suggests that it is not productive to have low ability students spend a long time experiencing negative affect in the face of failure. If so, the LE should give more direct hints and explanations. Another promising strategy to manage confusion is one recommended by VanLehn in his research on impasses during learning [16]. This strategy takes effect when confusion is detected and it entails: (a) prompting the student to reason and arrive at a solution, (b) prompting the student to explain their solution, and (c) providing the solution with an explanation only if the student fails to arrive at an answer. Further research will be required to compare the effectiveness of these interventions that aim to promote learning by inducing and intelligently managing confusion.

Acknowledgments. We thank our research colleagues in the Emotive Computing Group at the University of Memphis (http://emotion.autotutor.org). 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.

References

1. Csikszetmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990)
2. Dweck, C.: Messages that Motivate: How Praise Molds Students' Beliefs, Motivation, and Performance (In Surprising Ways). In: Aronson, J. (ed.) Improving Academic Achievement: Impact of Psychological Factors on Education, pp. 38--61. Academic Press, Orlando (2002)
3. Stein, N., Hernandez, M., Trabasso, T.: Advances in Modeling Emotions and Thought: The Importance of Developmental, Online, and Multilevel Analysis. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (eds.) Handbook of emotions, 3rd ed, pp. 574--586. Guilford Press, New York (2008)

4. Lepper, M., Woolverton, M.: The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In: Aronson, J. (ed.) Improving Academic Achievement: Impact of Psychological Factors on Education, pp. 135--158. Academic Press, Orlando (2002)
5. Meyer, D., Turner, J.: Re-Conceptualizing emotion and motivation to learn in classroom contexts. Educational Psychology Review, 18(4), 377--390 (2006)
6. Schultz, P., Pekrun, R. (eds.): Emotion in Education. Academic Press, San Diego (2007)
7. Immordino-Yang, M.H., Damasio, A.: We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. Mind, Brain and Education, 1(1), 3--10 (2007)
8. Arroyo, I., Woolf, B., Cooper, D., Burleson, W., Muldner, K., & Christopherson, R.: Emotion Sensors Go to School. In: Dimitrova, V., Mizoguchi, R., Du Boulay, B., Graesser, A. (eds.) Proceedings of 14th International Conference on Artificial Intelligence in Education, pp. 17--24. IOS Press, Amsterdam (2009)
9. Conati, C., Maclaren, H.: Empirically Building and Evaluating a Probabilistic Model of User Affect. User Modeling and User-Adapted Interaction, 19(3), 267--303 (2009)
10. Forbes-Riley, K., Litman, D.: Adapting to Student Uncertainty Improves Tutoring Dialogues. In: Dimitrova, V., Mizoguchi, R., Du Boulay, B., Graesser, A. (eds.) Proceedings of 14th International Conference on Artificial Intelligence in Education, pp. 33--40. IOS Press, Amsterdam (2009)
11. Robison, J., McQuiggan, S., Lester, J.: Evaluating the Consequences of Affective Feedback in Intelligent Tutoring Systems. In: Muhl, C., Heylen, D., Nijholt, A. (eds.) Proceedings of International Conference on Affective Computing & Intelligent Interaction, pp. 37--42. IEEE Computer Society Press, Los Alamitos (2009)
12. Graesser, A., Jeon, M., Dufty, D.: Agent Technologies designed to Facilitate Interactive Knowledge Construction. Discourse Processes, 45(4-5), 298--322 (2008)
13. Litman, D., Silliman, S.: ITSPOKE: An Intelligent Tutoring Spoken Dialogue System. Paper presented at the Human Language Technology Conference: 4th Meeting of the North American Chapter of the Association for Computational Linguistics, Boston, MA (2004)
14. Brown, J., VanLehn, K.: Repair Theory: A Generative Theory of Bugs in Procedural Skills. Cognitive Science, 4, 379--426 (1980)
15. Carroll, J., Kay, D.: Prompting, Feedback and Error Correction in the Design of a Scenario Machine. International Journal of Man-Machine Studies 28(1), 11--27 (1988)
16. VanLehn, K., Siler, S., Murray, C., Yamauchi, T., Baggett, W.: Why Do Only Some Events Cause Learning during Human Tutoring? Cognition and Instruction, 21(3), 209--249 (2003)
17. D'Mello, S., Graesser, A.: Inducing and Tracking Confusion and Cognitive Disequilibrium with Breakdown Scenarios. Memory and Cognition (in press)
18. Graesser, A., Chipman, P., King, B., McDaniel, B., D'Mello, S.: Emotions and Learning with AutoTutor. In: Luckin, R., Koedinger, K., Greer, J. (eds). 13th International Conference on Artificial Intelligence in Education, pp. 569--571. IOS Press, Amsterdam (2007)
19. Bjork, R.A., Linn, M.C.: The Science of Learning and the Learning of Science: Introducing Desirable Difficulties. American Psychological Society Observer, 19, 3 (2006)
20. Craik, F.I., Lockhart, R.S.: Levels of Processing: A Framework for Memory Research. J. of Verbal Learning & Verbal Behavior, 11(6), 671--684 (1972)
21. Graesser, A., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., Gholson, B.: Detection of Emotions during Learning with AutoTutor. Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada (2006)
22. Piaget, J.: The origins of intelligence. International University Press, New York (1952)
23. Vygotsky, L.: Mind in society: The development of higher psychological processes. Harvard University Press, Cambridge (1978)