Indeterminacy and Context Challenges in Automated Team Assessment and Tutoring

Wayne W. Zachary[1[0000-0001-5610-6777]]

[1] Starship Health Technologies, 2250 Hickory Road, #150, Plymouth Meeting, PA 19462, USA

Abstract. A key difference between individual tutoring and team tutoring is the degree of control that the individual has on the trajectory and outcome of the problem or process of interest. In the individual case, the tutee does not have to share control of the problem solving process with others, while in the team case each tutee has only partial control of the overall response to the problem being solved. This creates problems of indeterminacy for assessment and tutoring, as the prior actions (and the effects of those actions) become a context for the assessment of any given team member's decisions and actions at any point in the problem's evolution. Indeterminacy makes individual and whole-team assessment more difficult and creates new context-tracking requirements for team tutoring systems. Pedagogical and technological solutions from prior team trainers are reviewed, and outlines for general solutions are suggested for future team tutors.

Keywords: team tutoring systems, assessment, cognitive diagnosis, Indeterminacy, Advanced Embedded Training System, context tracking, recognition-based model assessment.

1 Introduction

The modern period of computation research into instruction began with Bloom's [1] seminal 1984 paper on human instruction, that showed a two-sigma increase in learning performance for individually-tutored students over those with traditional class-room based instruction. Bloom's result was associated with the insight that human tutors implicitly used experiential learning by basing tutoring on student's work in applying knowledge and skills in actual problems and tasks. Since then, the field has largely focused on understanding how individual tutors achieve that effect and how it could be replicated in Intelligent Tutoring Systems (ITSs). Over the last thirty years, ITS research has been applied to many domains [2-5] in an empirical process of using learning science to create new tutoring methods, and effectiveness assessments to identify and refine tutoring models that work best. This has resulted in a general theory of intelligent tutoring [6-10] that focuses on individualized assessment and individualized-assessment-driven scaffolding for learning.

At the heart of the ITS endeavor has been the dual problems of behavioral assessment and cognitively diagnostic assessment [11] given the behavioral assessment. The latter refers to the highly inferential process of assessing the cognitive processes and specifically the knowledge state of the learner in a way that diagnoses the state of the learner's expertise or mastery of the knowledge and cognitive skill involved. These two levels of assessment vary in their complexity based, to a large degree, on the characteristics of the underlying problem domain and skill being learned. In very well-structured tasks and domains involving a single person working alone, such as solving algebra problems, a given decision or action can always be immediately determined as either correct or incorrect from the problem state at the time of the action. Cognitive assessment can similarly be more easily done in such domains because the required knowledge and canonical problem-solving process can be precisely and unambiguously defined as a deductive process. This allows the problem-solving process to be diagnostically assessed in terms of its conformance with the deductive application of the declarative and procedural knowledge involved. Assessment of correct behaviors then leads to an increased belief that the learner has internalized and mastered the knowledge required for that particular problem step, and assessment of incorrect behaviors analogously leads to a decreased belief that the learner has internalized and mastered the underlying knowledge.[1]

There are of course many domains where the problem-solving processes are not so well structured. Many of these are discovery-based, or involve stochastic relationships between actions and outcomes. These are domains for which assessment of a behavior or action can be complex and/or can yield an indeterminate result. If behavior assessment is problematic, then cognitive assessment will be similarly problematic. The difficulties grow significantly greater when an ITS is trying to train individuals for problems in which the learner:

- is participating in an interaction with another person (who may be cooperating, coordinating, or even competing with the learner), or
- is part of a team of learners, either working alone or in collaboration, or in competition or conflict with each other.

In such interactive and team-based problem domains, the task of automated behavioral assessment quickly becomes very complex and problematic. As it does, the challenge of automated cognitively diagnostic assessment also becomes that much more difficult.

The remainder of this paper focuses on issues that underlie that difficulty – the issues of context-dependence and the problem of indeterminacy. These concepts, and the problems they create for team ITS, are discussed below within a detailed, though abstracted, example.

---

[1] This discussion is deliberately avoiding the mathematical and computational aspects of representing, increasing, and decreasing the belief that the learner has acquired specific elements of knowledge. An excellent presentation of those issues is provided in Nichols, Chipman, and Brennan [11].

2 Indeterminacy and Context-Dependence in Team Assessment

In a classical ITS, the learner is immersed into a practice environment in which, at each action or decision point, the:

- learner is in full control, and
- behavioral and cognitive assessment are done from direct observations of the state of the environment in plus the observed decision made or action taken.

This individual ITS model is a direct analog of one-on-one tutoring, as discussed by Bloom [1]. When this constraint is relaxed by adding other persons to the problem-solving process, it becomes more difficult to assess the actions of any one learner, and arguably impossible to do so using only direct observation of each actor in isolation. Consider a team-training ITS for the simple case of two persons in a simulated vehicle -- a pilot or driver and a navigator-communicator (navcom). Assume that the role of the navcom is to:

- a) plot a route to destination for the vehicle and communicate to the pilot the starting and ending point of the next segment;
- b) communicate with any external sources about problems or issues in the space to be crossed (e.g., locally bad weather); and
- c) revise the route and communicate changes to the pilot accordingly.

Assume further that the role of the pilot is then to direct the vehicle at all times, taking into account local conditions and other events or objects that may be relevant to safe operation of the vehicle.

In the physical world, it would not possible to assess the behavior of the pilot without considering the behavior of the navcom. If the navcom, for example, ignores information about an upcoming obstacle, and the pilot collides with it, then it is uncertain whether the pilot's behavior was correct or incorrect, making it generally impossible to assess that behavior or the knowledge state or cognitive process behind the pilot's action. The outcome was clearly negative (a crashed vehicle), but one could reasonably note that the pilot was just following the route provided by the navcom (Case 1). Or, one could determine that the pilot should have avoided the obstacle even without the navcom's inputs, as part of competent piloting skills (Case 2). Or, one could find that the pilot was deficient in being too dependent on the navcom's inputs, and not exercising normal caution that would be appropriate *if the pilot were in the vehicle alone,* performing both roles (Case 3). To add to this confusing picture, it should be noted that this assessment process is largely dependent, not on the prior standard of what the pilot or the navcom should do, but rather anchored on the way in which the coordination between the two roles was defined – that is, on *how the team interactions and coordination processes are defined*.

The above example points out why a team ITS cannot simply be viewed as an aggregate of individual ITSs for each member of the team. Examining this from the ITS architecture perspective, the behavioral and cognitive assessment of the ITS for Case 1 can only be accomplished by adding an independent (data) pipeline from the pilot's actions to the assessment module. However, it would be insufficient for Case 2, because a behavioral input alone would not allow the value of the missing communication from the navcom to be expressed and used in the behavioral and cognitive assessment of the pilot. Moreover, the addition of a pipeline from the behavior (and behavior assessment) of the navcom would still leave the assessment module for the pilot without enough information to consider and assess Case 3 above. That is because the information on the navcom's role in the team, vis-a-vis the pilot's role, would still be missing.

The point here is that knowing that the pilot drove into an obstacle leaves the behavioral (and cognitive assessment) in an *indeterminate* state with regard to diagnosis and assessment. In abstract terms, the arrival of the team-directed system at a specific problem-state can be the result of a (potentially large) set of unique sequences of actions/decisions by the members of the team. Different sequences in this set can be the basis for different diagnoses and assessments of some or all of team members at that same point in the problem state. In cases like this, we can say the there is an *indeterminacy* of the problem state with regard to diagnosis and assessment because a single diagnosis and assessment cannot be determined without additional information.

In the example immediately above, the additional information needed is historical -- the sequence of prior actions and interactions of and among the team members. However, other kinds of information may also be needed to undertake a definitive assessment and diagnosis. The individual action/decision sequences also involve the different relationships that the team members have to the set of roles and responsibilities that are defined within the team as a whole. The actions taken by individuals acting as a specific role can also have a situational meaning in terms of the changing state of the environment or situation that is the focus of the team. Together, the historical decision/action sequence of the different team members, and the relationship of actions/decisions to the team members to design, and the external problem state constitute a broader *context* for the assessment processes of the individuals in each role and of the team as a whole.

The added importance of context can be understood by adding one additional factor to the thought exercise above. Assume that the navcom had received multiple warnings of expected obstacles and had communicated each one to the pilot, though each expected obstacle communication proved to be a false alarm. The presence of multiple prior warnings, all false, is relevant context for the collision with the obstacle that was struck without a navcom warning. The prior false alarms could be interpreted as negatively affecting the vigilance of the pilot, and perhaps that of the navcom as well, leading to a slowed reaction time to the actual obstacle (Case 4). This case requires a context-based assessment and diagnostic process which involve both past events and external parties (i.e., whoever was issuing the warnings) as well as all of the factors required to assess Cases 1 through 3.

3 Dealing With Indeterminacy and Context-dependence in Team Training ITSs

The issues of indeterminacy and context-dependence as related to behavioral assessment and cognitive diagnosis in team ITSs were first addressed in one of the first team ITSs, the Advanced Embedded Training System (AETS) [12]. That system, and its initial solution to those challenges, are described in the following subsection. While AETS's approach created a foundation that continues to be relevant to today, it left other problems in team ITS design and development open. Some of those issues are also discussed in this section.

3.1 AETS and Recognition-Activated Model Assessment

The Air Defense team in the combat-information center (CIC) team aboard a US Naval destroyer focuses on the problem of commanding and controlling multiple assets to provide continuous defense of ownship and the whole surface combatant group from hostile attack from the air. The team can vary in size from six to eight members (within the CIC), with roles varying to some degree according to the mission and organizational decisions by the ship commander. The broad aid defense function is to detect, identify, monitor and, if necessary, engage air vehicles that could pose a threat to ownship and/or defended assets, particularly an aircraft carrier. The AETS was an advanced development research project that was undertaken to explore how adaptive intelligent training could be provided while at sea for whole shipboard teams, such as the Air Defense team.

(The initial motivation for AETS arose out of a specific incident that occurred in the late 1980s, in which a US Naval destroyer shot down an Iranian airliner with great loss of life. The Air Defense team believed, based on the aircraft's unusual behavior and the high level of geopolitical tensions in the area, that the aircraft was in fact a hostile military aircraft preparing to launch a missile at the destroyer. This incident was widely analyzed in a landmark study on decision-making under stress [13], which essentially concluded that all the actions of the team were appropriate, although contextual factors led to the clearly undesirable outcome, making it an interesting empirical example of the issues addressed in this paper.)

AETS initially focused on applying conventional ITS concepts, seeking to assess each team member's performance from bottom-up analysis of that person's low-level actions -- specific keystrokes, eye movements, and speech utterances made by the operator[2] on the voice networks. It quickly became clear that there were very many sequences of low-level actions that could be used to create a functional event in the problem solving process, e.g., tagging an air track as presumed hostile. Cognitive front-end analyses [14] also showed that those abstracted functional events were the basis on which operators, particularly those in more senior roles, reasoned about the problem. The cognitive analyses also showed that each operator maintained a detailed mental model of the mission context from the perspective of that operator's specific role in the

[2] The term 'operator' is used henceforth to refer to a person filling a specific role in the team.

team, and used that mental context model to stimulate opportunistic reasoning about what to do next. This reasoning strategy stood in stark contrast to the top-down deductive reasoning described earlier as the canonical individual ITS case. One seeming basis for this use of the opportunistic context-driven reasoning approach was that it was an implicit response to the indeterminacy in the team process. As each team member could independently move the problem in an unexpected direction (i.e., could create indeterminacy), the experienced operators developed a strategy that explicitly maintained a context representation, and that at any point in time reacted to the situation at hand, in the context of the current mission.

The AETS behavioral and cognitive diagnosis approach mimicked the strategy uncovered in the cognitive analysis. It consisted of four parts:

1) low-level action data were processed using intelligent algorithms to automatically combine them into abstracted high-level actions, which marked the key steps and transitions in the problem-solving process;
2) cognitive models were constructed to emulate the processes by which each operator role built and maintained a mental model of the mission context, and the processes by which the operator chose (and contextualized) high-level actions to take;
3) performance analysis algorithms, on recognition of a high-level action from an operator, queried the cognitive model to determine if its type, timing, and contextual customization matched the high-level action, if any, that were indicated by that operator's cognitive model;
4) cognitive analysis algorithms were then invoked, given an at-least partial match with the model indications, to identify the specific parts of the cognitive model that were successfully or unsuccessfully instantiated in the operator's actions; and
5) adaptive feedback algorithms then used the results of the cognitive assessment to provide feedback reinforcing the knowledge used correctly or attempting to remediate inferred errors in underlying knowledge.

This process was termed Recognition Activated Model Assessment (RAMA), since it was activated by recognition of an abstract functional action from an operator, and conducted through comparison of the action with underlying cognitive model predictions. The performance assessment subsystem also used temporal windowing to control for small variations in timing of actions by operators, and to allow missed actions to be recognized by their absence. This RAMA approach has been used in various forms by multiple other team training ITSs [15]. Among other novel features of RAMA was its use of explicit context models (though for each individual operator rather than one team-wide), and its use of abstract levels of action to drive the assessment process rather than unitary or low-level actions typical of conventional ITSs.

3.2 The Inverse Indeterminacy Problem – Creating an Assessable Moment

AETS and the RAMA method still left several indeterminacy problems unaddressed. Among the most interesting was a way of meeting a training need that can be considered the inverse of the indeterminacy problem. That was the problem of creating an assessable moment: a specific situation that required one or more operators to demonstrate their possession and ability to apply a specific body of knowledge. In an individual ITS this is relatively easy; a problem or state can be created directly by the ITS designer or engineered so that the learner must encounter it. In a team environment, however, it is much more difficult to do this for the very reason underlying indeterminacy, which is that each and any operator could move the problem in some unanticipated direction. Thus, creating a specific situation requires that each operator behave in such a way as make that situation arise, or at least require that no operator behave in a way that would prevent the situation from occurring.

A successor to AETS called SCOTT (Synthetic Cognition for Operational Team Training) did explicitly address the problem of creating assessable moments from within a RAMA architecture [16]. It did this explicitly by creating a team training ITS in which any role in a team can be trained, but in which only one role is played by a live human trainee at a time, with the other roles being filed by cognitive models interacting directly with the simulation. Thus, a SCOTT cognitive model served dual purposes: as the basis for RAMA assessment when its role was being played by a live trainee, and as a synthetic operator otherwise. In doing this, SCOTT was designed so that the model-based operators could be directed to secretly collaborate to create an assessable moment for the live trainee.

4 Summary and Future Directions

This paper has discussed several challenges to the task of constructing intelligent training systems for teams, as follows:

- In moving from the classical paradigm of one-learner/one-ITS to the team-training paradigm of many-leaners/one-team-ITS, some or all of the teammates become part of the problem environment for assessing the behavior and knowledge state for any individual in the team.
- Because the design and standard procedures for the team roles affect how any individual action is assessed, the team and its design also become part of the problem environment for assessing the behavior and knowledge state of any team member. The team level also creates a separate level of assessment for the team as a whole.
- The history of team members' actions and the effect of those actions on the external problem environment create a persistent context that also provides needed information to the individual and team level performance assessment process. Within the team, different members may have differential access to this larger team context information.

Creating explicit representations of these additional second and third order influences on individual team-member assessment and diagnosis will be required in future team ITSs to provide tutoring for team interactions and cooperation and coordination within a team. The paragraphs below speculate on how this might be accomplished.

The prior generations of intelligent team trainers (see Freeman and Zachary [15]) relied to various degrees on human instructors, role-players and/or observers as adjust of the otherwise automated team trainers. In AETS, for example, the human instructors were responsible for tracking team communications and collecting specific examples of those communications to use in live after actions reviews with the (human) team. While the embedded cognitive models in AETS did build and maintain cognitive representations (termed mental models) of the team and problem context, each such model only considered it from the perspective of one specific role/person in the team. Those computational context representations provided the information needed by the position-specific RAMA algorithms. There were two main limitations of this approach. The first is that there was no model of the overall 'team' context, so problems and failures that resulted in divergent context representations within the team could never be detected or diagnosed. Second, individual context view is insufficient to represent coordinated or cooperative aspects of teamwork, again preventing such aspects from being assessed or diagnosed. These limitations require explicit models of context and/or of team communications to be developed and integrated into the (simulated) practice environment.

One emerging technology that could be used to accomplish this is computational context modeling[3], a family of technologies that seek to build and maintain dynamic declarative computational models of context. Particularly relevant for team ITSs are context-modeling approaches that seek to construct a representation that is compatible with the mental models of context models that people construct [18]. A drawback of this approach is that it can require intensive knowledge-engineering, especially for larger teams. An attractive aspect, on the other hand, is that a representation of the 'core' context that is shared across the team can be constructed, and more specialized role-specific context models (analogous to those that were used in AETS and SCOTT) can be generated easily using a publish-subscribe mechanism augmented with more detailed context information maintained separately. This context mechanism can also be used to maintain a context-based history of communications and dialogs among team members. Zachary, Carpenter, and Santarelli [19] detail an example of this from a human-robot communication domain.

The complexity of such a thorough context representation technology could make it prohibitively expensive if it had to be (re-)built from scratch for each new team ITS. However, a substantial economy of scale could be achieved by integrating it as common infrastructure in re-usable team ITS framework. A team-focused GIFT [20] could thus provide a logical insertion point for this key component of future team ITSs.

References

1. Bloom, B.S. The 2-sigma problem: The search for methods of group instruction as effective as one- to-one tutoring. Educational Researcher, **13**, 4-16. (1984).
2. Nwana, H. S. Intelligent tutoring systems: an overview. Artificial Intelligence Review, **4(4),** 251-277. (1990).

[3] See, for example, the various papers in Lawless, Mittu, Sofge, & Morrison [17].

3. Frize, M., & Frasson, C. Decision-support and intelligent tutoring systems in medical education. Clinical and investigative medicine**, 23(4)**, 266-269. (2000).
4. Nkambou, R, Mizoguchi, R., & Bourdeau, J. (Eds.). Advances in intelligent tutoring systems (Vol. 308). Springer Science & Business Media. (2010).
5. Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H., & Gilbert, S.B. Designing adaptive instruction for teams: A meta-analysis. International Journal of Artificial Intelligence in Education, 1-40. (2017)
6. Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. Intelligent tutoring systems with conversational dialogue. AI magazine, **22(4)**, 39. (2001).
7. Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. International Journal of Artificial Intelligence and Education, **15**, 261-290. (2005).
8. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, **46(4)**, 197-221. (2011).
9. Kulik, J. A., & Fletcher, J. D. Effectiveness of intelligent tutoring systems: a meta-analytic review. Review of Educational Research, **86(1)**, 42-78. (2016).
10. VanLehn, K. Regulative loops, step loops and task loops. International Journal of Artificial Intelligence in Education, **26(1),** 107-112 (2016).
11. Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). Cognitively diagnostic assessment. Routledge. (2012).
12. Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. The Advanced Embedded Training System (AETS): An intelligent embedded tutoring system for tactical team training. International Journal of Artificial Intelligence in Education, **10**, 257-277. (1999).
13. Cannon-Bowers, J.A., & Salas E. (Eds.). Making decisions under stress: Implications for individual and team training. Washington, DC: APA, (pp. 271-297). (1998).
14. Zachary, W., Ryder, J., and Hicinbothom, H. Building Cognitive Task Analyses and Models of a Decision-Making Team in a Complex Real-Time Environment. In Schraagen, J.M.C., Chipman, S.F., & Shalin, V.L. (Eds.). Cognitive Task Analysis. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. pp. 365-383. (2000).
15. Freeman, J., and Zachary, W. Intelligent Tutoring for Military Team Training: Lessons learned from US Military Research. In Johnson, J., Sottilare, R., Sinatra, A., Burke, S (Eds). Building Intelligent Tutoring Systems for Teams: What Matters. Bingley, UK: Emerald Publishing. (2018).
16. Zachary, W., Scolaro, J., Stokes, J., Weiland, W., & Santarelli, T. Using synthetic naturalistic worlds to train teamwork and cooperation. Scaled Worlds: Development, Validation, and Applications, 316. (2004).
17. Lawless, W., Mittu, R., Sofge, D, and Morrison, J. (Eds.) Computational context: The value, theory and application of context with AI. Boca Raton, FL:CRC Press. (2018).
18. Zachary, W., Rosoff, A., Miller, L., and Read, S.. Context as Cognitive Process: An integrative Framework for Supporting Decision Making. In Laskey, K., Emmons, I., and Costa, P. (Eds). Proceedings of 2013 Semantic Technologies in Intelligence, Defense and Security. CEUR Conf Proceeding Vol-1097. Pp48-55 http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-1097/STIDS2013_T07_ZacharyEtAl.pdf (2013).
19. Zachary, W., and Carpenter, T. Using context and robot-human communication to resolve unexpected situational conflicts. In 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). Pp.1 - 7, DOI: 10.1109/COGSIMA.2017.7929596 (2017).

20. Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., ... & Winer, E.. Creating a team tutor using GIFT. International Journal of Artificial Intelligence in Education, 1-28. doi:10.1007/s40593-017-0151-2. (2017).