Enhancing Human Understanding through Intelligent Explanations

Tina Mioch[1,2], Maaike Harbers[1,2], Willem A. van Doesburg[2], and Karel van den Bosch[2]

[1] Institute of Information and Computing Sciences, Utrecht University, P.O.Box 80.089, 3508 TB Utrecht, The Netherlands
[2] TNO Defence, Security & Safety, Kampweg 5, 3796 DE Soesterberg, The Netherlands
{tina.mioch,maaike.harbers,
willem.vandoesburg,karel.vandenbosch}@tno.nl
http://www.tno.nl

Abstract. Ambient systems that explain their actions promote the user's understanding as they give the user more insight in the effects of their behavior on the environment. In order to provide individualized intelligent explanations, we need not only to evaluate a user's observable behavior, but we also need to make sense of the underlying beliefs, intentions and strategies. In this paper we argue for the need of intelligent explanations, identify the requirements of such explanations, propose a method to achieve generation of intelligent explanations, and report on a prototype in the training of naval situation assessment and decision making. We discuss the implications of intelligent explanations in training and set the agenda for future research.

Key words: Explanations, simulation-based training, intelligent tutoring systems, cognitive modeling, feedback, learning

1 Introduction

Human well-being and performance are highly affected by the environment in which a person operates. People are always trying to improve their conditions, from increasing the temperature when it is cold to developing more and more advanced computer systems to aid them in their daily work. A recent development in the enhancement of environments is the incorporation of mechanisms that show some understanding of humans. Such mechanisms use sensors to acquire information about human functioning and analyze this information to adapt to human needs. An environment containing systems with these mechanisms is called intelligent.

Some of the applications exposing such ambient intelligence require interaction between the human user and the system. For example, decision-support systems have to communicate their advice to the person who is in charge of making a decision, and tutoring agents need to convey instructions and feedback to

a student. Human-system interaction has two sides: the system or agent has to transmit information to the human, but it also has to understand the human. An agent reminding elderly or disabled people to take their medicine not only has to convey this message, it must also be able to understand when someone says he or she has already taken the medicine. One of the requirements of good human-system interaction is that the human understands and accepts a system's message. The quality of interaction between the human and the system is an important factor in the endeavor to improve human comfort and performance.

We claim that one of the factors contributing to the quality of human-system interaction is intelligent explanation. Providing explanations along with presented information is not something new. Various explanation components have been developed in recent decades for software systems, such as intelligent tutoring systems, decision-support systems and expert systems [8, 4, 9, 12]. It is supposed that the more these explanations are tailored to the specific needs of the user, the better the user is served. A system could make distinctions between users on the basis of their knowledge, speed of learning, most efficient learning method, preferences, etc. Most of the existing explaining components do not take features of the specific user into account, but treat all users in the same way.

In this paper we will clarify that in order to improve the effectiveness of explanations, systems should be equipped with capacities that refer to the users' mistakes, performance, beliefs, knowledge, intents and the like in their explanations. First we will take a closer look at agent explanation in ambient systems: what are the requirements to make an explanation useful, and what type of explanations can be distinguished? Then we will discuss how such an explanation mechanism can be implemented in a feedback system of a simulation-based training environment.

2 Intelligent Explanations

Most people take the information on a digital clock for granted; there is no need for further explanation about the current time. However, a user is not always sufficiently informed by such basic information. Even though a system may be correct in stating 'It is time to buy a new computer', this announcement might raise some questions. He or she wants to know why the computer believes this; is the computer too outdated for its purpose, or is the computer broken? If so, it would be interesting to know what part is not functioning and whether there are possibilities to update or repair the computer. This example shows that in some cases it is not sufficient if a system just presents its conclusion. An accompanying useful explanation will make more sense to the human user and it will increase both the human's understanding and acceptance of the system [20, 21].

Explanations exist in many forms. Furthermore, one single event can be explained in different ways. One explanation is not by definition better then another; the desired explanation depends on the context in which it is given. For example, a possible answer to the question 'Why did the apple fall?' is 'Because I dropped it'. In some situations however, the explanations 'Because I stumbled'

or 'Because he pushed me' would be more useful. A whole other type of explanation of why the apple fell is 'Because of the gravitation force'. Dependent on the context, people need a particular type of explanation. An explanation system should be able to estimate the information need of the user and provide an explanation accordingly.

Another difficulty to overcome in providing explanations is *timing* [20]. In some applications it is obvious that each time new information is presented it should be accompanied by an explanation. For instance in diagnosis systems, every given diagnosis should be accompanied by an explanation of how the system came to this result. In contrast, in systems that constantly provide new information, there are no predefined moments in which explanations should be given. For instance, a navigation system has to decide for itself when the user needs new instructions. So in a complex and open environment, an explanation system should be able to determine when and how often the user needs explanations.

Furthermore it is desirable that explanations are adapted to the receiver, as not all people are the same and thus might need different explanations. Whereas novices tend to need extensive explanations, experts generally prefer explanations in which the to them obvious steps are skipped [17]. Besides level of expertise, other human factors such as knowledge, intents and emotions could be taken into account. An explanation commenting on an assumed strategy of a student could be: 'Because you performed action a_1, I think your plan must be P. This is not a good strategy because you do not have enough resources to perform action a_3, which is also part of plan P'. An explanation involving emotions is: 'The other agents acted this way, because your angry words scared them'. Hence, intelligent explanations should be adapted to the user's perspective to enhance understanding and learning.

3 Related Work

In the past twenty years, much research has been done on intelligent tutoring systems (ITS) [14, 13, 4], which are systems that teach students how to solve a problem or execute a task by giving explanations. Such systems have been successfully designed for the training of well-structured skills and tasks (e.g. LISP programming [11] or algebra [10]), which are relatively closed, involve little indeterminacy, and do not involve real-time planning. For the training of real world tasks, these conditions do not always apply [1]. Real world tasks are often complex, dynamic and open in the sense that outcomes of actions may be unpredictable. These features make it difficult to design training, because it is usually not possible to represent the domain by a small number of rules. Moreover, the space of possible actions is large. For instance, the military uses simulation-based training systems to train tactical command and control [2]. In such training, the student responds in real-time to simulated problems, so the system needs to be able to evaluate whether the actions taken are correct and whether they have been executed at the right time. A complicating characteristic of evaluating tactical performance is that there is often no single 'right' way to

accomplish a task, but that there exists more than one good solution for a problem, depending on the context [3]. In addition, a training system should not only evaluate a student's behavior, but, in case of errors, it should also take cognitive processes underlying that behavior into account. The result should be suitable for inferring the student's strategy. The demands on context-sensitivity and performance diagnosis make it hard to generate appropriate feedback.

In recent years, the challenge of developing and providing explanations in open, complex and dynamic environments has been accepted by the international research community [7, 15, 5] and first steps forward have been made. Livak, Heffernan and Moyer [7] developed a prototype that uses a cognitive model to provide both tutoring and computer generated forces capabilities. The actions of the student are evaluated by comparing the student's behavior to the ideal behavior of an expert. If the student deviates from the behavior that the expert model demonstrates, feedback is returned to the student. The feedback that is given is a low-level explanation of why the particular action at that moment was not correct. The explanations only refer to a particular action, and no reference to consequences of actions are given. Furthermore, the tutoring agent does not maintain a model of the user to take his beliefs and intentions into account. As a consequence, the tutoring agent is not capable of adjusting its feedback to the specific knowledge and intentions of the student.

Other research focuses on the debriefing phase of the training by letting the simulation entities explain their reasons for executing particular actions to the student. Examples of this approach are the explanation system *Debrief* [8] and the *XAI system* [9]. Debrief is used to generate explanations for the individual agent's actions in the debriefing phase of the simulation, together with information about what factors were critical for taking that action. The XAI System allows the student to further investigate what happened during the exercise. In order to generate explanations, the software agents log important actions annotated with abstract information about underlying reasons of the actions as well as their consequences. Both Debrief and XAI explain the reasoning behind the executed actions on demand, expecting the student to ask the relevant questions. No assessment of the student is made, so no directed feedback can be given to the student. In addition, the explanations are directly related to knowledge about the task, giving a low-level reason for a particular action. For example, for the task of clearing a room, an agent might answer the question 'Why did you throw a grenade into a room?' by stating that 'A grenade suppresses enemies that are in the room'. It would be more informative for the student to give an explanation on a higher conceptual level, including e.g. beliefs and intentions of the agent. Such an explanation would for example be 'I believed that the enemy was in the room. My goal is to clear all rooms. By throwing grenades into the room, I intended to suppress the enemy'.

As can be seen, research on explanations has been recognized as being important in training simulations to enhance the student's learning experience. However, a lot of research is still required. Questions that still need answering are for example how to obtain insight into the cognitive processes of the student,

and how to support students in acquiring an understanding of the relationships between their behavior and the consequences in the environment. To achieve understanding, explanations must be given about processes in the environment. However, as stated above, explanations in simulation-based training systems are often not profound enough to achieve this result.

4 Types of Feedback

When building simulation-based training systems, three types of feedback can be distinguished. They differ in the types of information that they take into account and the sophistication of explanations they give:

Result-based feedback: Feedback is based only on observable student behavior. Correct results, formulated by domain experts, are hard-coded into the scenario, and feedback is generated by comparing the student behavior with the correct behavior. The feedback states only whether the student has completed the task successfully, and if not, what the correct behavior should have been.

Model-based feedback: Feedback is not only based on explicit student behavior, but also on contextual knowledge of the simulation environment and explicit task knowledge. Using the different kinds of information, the feedback is generated by reasoning about an internal model.

Cognition-based feedback: As with model-based feedback, feedback is based on explicit student behavior, knowledge about the simulation environment and task knowledge. In addition, a user model is developed that makes it possible to infer cognitive strategies of the student to facilitate even better feedback.

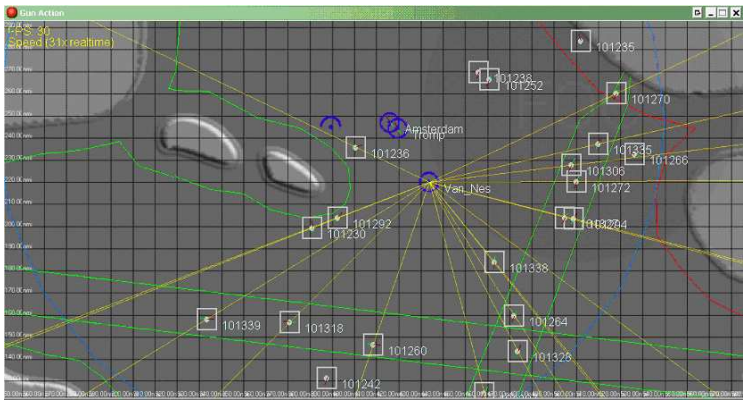

Fig. 1. Screenshot of our training scenario. The squares indicate radar tracks, the circles represent tracks of own forces.

We will illustrate how the types of feedback differ from each other in the context of a navy task, namely the tactical picture compilation task. Developing a tactical picture of the environment is an essential part of any military mission. In the tactical picture compilation task, the tracks in a radar picture have to be classified into categories such as the type of vessel, and the probable intention of the vessels has to be determined. An illustration of a situation in which a student has to develop a tactical picture can be seen in Fig. 1. Assuming that the student has to decide at a particular point in time which track poses the largest threat to the ship, the following examples of explanations illustrate the types of feedback that the system might give.

Result-based explanation *Your answer is incorrect: You have chosen track 101304 as the most threatening track. It should have been track 101112.*

Model-based explanation *The expert disagrees with your answer: You have chosen track 101304 as the most threatening track. However, its speed is not very high. Additionally, there are a number of ships that are moving at the same speed AND are closer to you. The expert thinks the most threatening track is 101112.*

Cognition-based explanation *The expert disagrees with your answer: You have chosen track 101304 as the most threatening track. However, its speed is not very high. Additionally, there are a number of ships that are moving at the same speed AND are closer to you. The expert considers 101112 to be the most threatening track.*

You have assessed tracks coming from the harbors, probably because you might suspect the enemy to reside in the harbor. However, a good strategy is to start investigating close to your vessel and progress outward. We have not seen you do this in the scenarios you have played thus far. This tip should help you protect your ship by preventing enemy vessels to get too close unseen.

These examples illustrate that generating model-based explanations are the minimum requirement for *intelligent* explanations. Cognition-based explanations are the most sophisticated and would be the type of explanations from which a student learns the most. For that reason, our goal is to build a training system that generates cognition-based explanations.

5 Intelligent Tutoring Agent for the Royal Netherlands Navy

For the Royal Netherlands Navy we investigate the possibility of developing an agent that fulfills the task of an instructor in a training simulation. We focus on the functionality of evaluating student performance and deliver this evaluation along with an explanation. The task that is chosen is a modification of the tactical picture compilation task as described in Sect. 4. In the modified task, the student is presented with a radar picture at a particular point in time, showing a number of radar tracks. The student has to gather and integrate information on

these tracks to form a mental tactical picture of the situation. Then the student has to decide which track poses the largest threat to his own ship. Time does not play a role as the picture is static and represents a situation at a particular point in time. Factors in this task that have to be taken into account are for example the speed of vessels, distance from own ship, whether they adhere to shipping lanes and whether they are inbound.

We are developing a training system that uses cognition-based explanations. To meet this objective, the following method is introduced. To generate feedback, an expert agent is executed, as are agents that deviate in some aspect from the expert, representing typical mistakes of students. These deficient agents intentionally fail to take one or more particular factors of the task environment into account, or are deficient in another way. It is assumed that errors of the student are the result of incorrect beliefs or an incorrect strategy. The assessment the student makes of the situation after examining the screenshot is compared to the assessment of the expert and the deficient agents. Four outcomes of the comparison can be differentiated.

First, the assessment of the student might not correspond to either the expert's assessment or to any of the deficient agents' assessments. In that case, the student did not complete the task satisfactorily nor did he make a typical mistake represented by any of the deficient agents. An explanation is then generated that explains why the assessment of the expert is preferable to the assessment of the student based on the comparison between the two performances and on task knowledge. This includes for example knowledge about the environment and knowledge about the importance of different relevant factors.

Second, the assessment of the student corresponds to the assessment of the expert agent, and to none of the assessments of the deficient agents. In this case, it is assumed that the student has solved the task satisfactorily, and accordingly, positive feedback is returned. As the environment is open, it is of course possible that the student's assessment to the problem is correct, but that the student has just been lucky, without the assessment being based on correct beliefs and a correct process of obtaining the assessment. However, over several trials of the training simulation, the incorrect strategy of the student will eventually fail and the student will then receive a feedback that shows that his beliefs are wrong.

Third, the student's assessment might correspond to the assessment of one particular deficient agent, without matching the expert agent or any other deficient agent. As it is assumed that the deficiency of the agent corresponds to the beliefs or strategies of the student, a diagnosis of the student's state of mind can be made and an explanation be generated. The explanation that is returned corresponds to the deficiency of the agent.

Fourth, it is possible that the student's assessment corresponds to several assessments, either of several deficient agents, or one or more deficient agents and the expert agent. This is possible because there are often many possible ways to arrive at the same assessment. Then, the response alone is not sufficient for deciding what feedback is appropriate. We need information about the processes that resulted in the selection of that response and which beliefs and strategies

the student used to obtain his response. If we can do this validly, then we can return feedback containing an appropriate explanation. In this case, the user model is of importance, because it gives extra background information about the process that led to the assessment. On the basis of the inferred beliefs and strategies of the student, it is possible to choose the most corresponding of the matching assessments and return the appropriate explanation.

Our prototype does not yet take performance over time into account. In reality, the history of the situation should be used in situation assessment. For example, an apparently non threatening radar track (taking only properties such as, speed, distance, bearing and adherence to shipping lanes into account) may in fact be highly suspect because it has recently varied its speed and has intermittently crossed the shipping lane. A student that is sensitive to this information correctly assesses this track as threatening. A system that cannot use such information may then, erroneously, 'correct' the student and thus fault the student for his judgment even though the student actually outperforms the expert model. Such problems are typical for evaluating performance in complex, dynamic and open tasks. To overcome these problems, it is more useful not to evaluate the assessments of a student but the cognitive strategies that have led to the assessment.

A problem is that cognitive strategies are not observable. We are faced with the problem to construct a user model containing hypotheses about the strategies of the student without the ability to observe these strategies. We choose to overcome this problem by arranging the training simulation in such a way that the user is forced to provide some information about his strategy. We achieve this by allowing the student access to all information that is available in the actual operational environment, but only on explicit request of the student. For example, by initially hiding the shipping lanes and allowing these to be seen for a short period of time we gain evidence that the student is checking adherence to shipping lanes. By observing the pattern of behavior while the student is executing the task, we can build a hypothesis of the strategy that the student employs. Moreover, we can test the hypothesis by selecting a subsequent scenario and predicting the steps that the student will take. If the hypothesis is confirmed, we can then confidently proceed in providing feedback on the strategy level rather than on the performance level. In addition, it enables us to select those scenarios that practice the particular aspect which the student finds difficult. Because the user model is taken into account, the feedback is based on the perceived process of decision-making of the student, which includes an interpretation of the student's actions. By giving an explanation that has relevance to the student's actions and underlying beliefs, acceptance and understanding of the feedback are endorsed.

6 Discussion

In this paper we argued for the importance of intelligent explanations in human-system interaction. We clarified why explanations should be user-specific and what aspects should be taken into account in order to achieve this. There are

different ways to generate model-based or cognition-based feedback; we use a method of modeling the user, an expert agent and deficient agents. The behavior of the user is compared to that of the agents. We argue that the results of such comparisons in combination with the user model yield insights about the user which make it possible to provide explanations fit to the particular user.

As mentioned before, we are currently implementing our method of explanation generation in a training system for the Royal Netherlands Navy. Once the system is ready, we will evaluate whether the explanations generated by the proposed method will improve the users' performances. We aim to extend the method to other situations and to apply it to more complex versions of the task, for example involving a time component, and to other tasks than tactical picture compilation. Therefore the expert model, the user model and the deficient agents need to be adapted to the new demands. Despite these changes and modifications, the core mechanism of the approach remains the same. So if the explanations are satisfying in the simple case, we are confident that the system is able to generate intelligent explanations in more complex versions of the task and other tasks as well.

A system providing the desired intelligent explanations referring to knowledge, plans, intentions and the like will yield many advantages. First, good explanations will help the user in his or her learning process, because they will improve conceptual understanding [18]. Besides this advantage while learning the task, good explanations prolong the duration of an acquired skill [16]. A practical advantage is the reduction of training costs. When systems are able to generate human-like explanations, fewer instructors are needed to complete training and this will save costs. The usual weighing between costs and quality no longer has to be made. Finally, because students are no longer dependent on the presence of a trainer, they are more flexible and can train a task whenever they want.

In future research, it could be investigated how expert and deficient agents can be modeled more efficiently. Especially a practical way of modeling deficient agents is useful, because in complex tasks many of them are needed. For this, the behavior of real students can be used. Also the relation between different deficiencies could be examined: what behavior does a user with a combination of different deficiencies show and how is this reflected in the modeled deficient agents? Further, more attention could be paid to the presentation of explanations: which way of presenting leads to the highest performance? It could also be investigated for what type of tasks the intelligent explanations turn out to be the most useful.

Acknowledgements This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie) and by the research programs "Cognitive Modelling" (V524) and "Integrated Training and Instruction" (V406), funded by the Netherlands defence organisation.

References

1. 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 (1999) 257-277
2. Doesburg, W.A. van, Bosch, K., van den : Cognitive Model Supported Tactical Training Simulation. Proceedings of the Fourteenth Conference on Behavior Representation in Modeling and Simulation (BRIMS), Universal City, CA (2005) 313-319
3. Hutchins, S.G., Kemple, W.G., Porter, G.R., Sovereign, M.G.: Evaluating Human Performance in Command and Control Environments. Proceedings of the 1999 Command and Control Research and Technology Symposium, Newport, RI (1999) 50-52
4. Murray, T.: Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. International Journal of Artificial Intelligence in Education **10** (1999) 98-129
5. Ntuen, C.A., Balogun, O., Boyle, E., Turner, A.: Supporting Command and Control Training Functions in the Emergency Management Domain Using Cognitive Systems Engineering. Ergonomics **49** (2006) 1415-1436
6. Murray, W.R.: Intelligent Tutoring Systems for Commercial Games: The Virtual Combat Training Center Tutor and Simulation. AIIDE (2006) 66-71
7. Livak, T., Heffernan, N.T., Moyer, D.: Using Cognitive Models for Computer Generated Forces and Human Tutoring. 13th Annual Conference on Behavior Representation in Modeling and Simulation(BRIMS), Simulation Interoperability Standards Organization, Arlington, VA. (2004)
8. Lewis Johnson, W.: Agents that Learn to Explain Themselves. Proceedings of the Twelfth National Conference on Artifcial Intelligence, Seattle, Washington, United States, American Association for Artificial Intelligence **2** (1994) 1257-1263
9. Gomboc, D., Solomon, S., Core, M.G., Lane, H.C., van Lent, M.: Design Recommendations to Support Automated Explanation and Tutoring. Proceedings of the Fourteenth Conference on Behavior Representation in Modeling and Simulation (2005)
10. Koedinger, K.R., Anderson, J.R.: Illustrating Principled Design: The Early Evolution of a Cognitive Tutor for Algebra Symbolization. Interactive Learning Environments **5** (1998) 161-180
11. Anderson, J.R., Conrad, F.G., Corbett, A.T.: Skill Acquisition and the LISP Tutor. Cognitive Science **13** (1989) 467-506
12. Core, M.G., Traum, T., Lane, H.C., Swartout, W., Gratch, J., van Lent, M.: Teaching Negotiation Skills Through Practice and Reflection with Virtual Humans. Simulation **82**:11 (2006) 685-701
13. Psotka, J., Massey, D., Mutter, S. (eds.): Intelligent Tutoring Systems: Lessons Learned. Hillsdale, NJ: Lawrence Erlbaum Associates (1988)
14. Polson, M.C., Richardson, J.J.: Foundations of Intelligent Tutoring Systems. Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1988)
15. Jensen, R., Nolan, M., Chen, D.Y.: Automatic Causal Explanation Analysis in Combined Arms Training AAR. Proceedings of the Interservice / Industry Training, Simulation, and Educational Conference (I/ITSEC) (2005)
16. Bosch, K. van den: Durable Competence in Procedural Tasks Through Appropriate Instruction and Training. In: Harris, D. (eds.): Engineering Psychology and Cognitive Ergonomics, Aldershot, Ashgate **3** (1999) 431-438
17. Ramberg, R.: Construing and Testing Explanations in a Complex Domain. Computers in Human Behavior **12**:1 (1996) 29-47

18. Teichert, M.A., Stacy, A.M.: Promoting Understanding of Chemical Bonding and Spontaneity through Student Explanation and Integration of Ideas. Journal of Research in Science Teaching **39**:6 (2002) 464-496
19. Morrison, P., Barlow M.: Child's Play? Coercing a COTS Game into a Military Experimentation Tool. Proceedings of SimTecT 2004, Canberra (2004) 72-77
20. Nakatsu, R.T.: Explanatory Power of Intelligent Systems: A Research Framework. Proceedings of the IFIP International Conference on Decision Support Systems, Prato, Italy (2004) 568-577
21. Herlocker, J.L.: Position Statement - Explanations in Recommender Systems. Proceedings of the CHI' 99 Workshop, Pittsburgh, USA (1999)