Error-Associated Behaviors and Error Rates for Robotic Geology

Geb Thomas, Jacob Wagner and Justin Glasgow

Department of Mechanical and Industrial Engineering

The University of Iowa

Robert C. Anderson

Jet Propulsion Laboratory, California Institute of Technology

Submitted to the Annual Conference of the Human Factors and Ergonomics Society

Introduction

When complex systems fail, people want to know the source of the failure. Humans are blamed for anywhere between 20 % (Richei, Hauptmanns, & Unger, 2001; Sugarman, 1979) and 80 % (Danaher, 1980; Lee, Tillman, & Higgins, 1988) of system failures, depending on the definition of error chosen for the study. Unfortunately, there is no consensus among the major researchers in the area regarding the best definition for human error (Hollnagel, 1993; Norman, 1993; Woods, J., Cook, & Sarter, 1994). A common definition of human error is a decision which leads to an unplanned, undesired, or generally bad result. In practice, such errors can decrease overall system effectiveness or even lead to system failure.

To more effectively design the robot's information systems for extraterrestrial geology, we would like to know what aspects of the system, the robot, the operator interface, or the geologist decision-making process, are the most problematic. Once the troublesome features are identified, they may be targeted for improvement, either by redesigning the robot, the interface, or training the geologists. In this work, we consider the challenge of understanding how geologist decisions lead to errors, based on an analysis of transcripts collected during a mock robotic field test conducted in October 2003 in the desert near Grey Mountain, Arizona.

This study explores human error as a function of the decision-making process. One of many models for human decision-making is Rasmussen's decision ladder [9]. The decision ladder identifies the multiple tasks and states of knowledge involved in decision-making. The tasks and states of knowledge can be classified by the level of cognitive effort required to make the decision, leading to the skill, rule, and knowledge taxonomy (Rasmussen, 1987). Skill based decisions require the least cognitive effort and knowledge based decisions require the greatest cognitive effort. Errors can occur at any of the cognitive levels.

Reason compiled sub-categories of errors in each of level of the taxonomy (Reason, 1990). Many laboratory and observational experiments, designed to identify specific behaviors leading to errors, have validated these classifications. In this work, we generalize the errors developed by Reason into eleven error-associated behavior characteristics: selectivity, workspace limitation, out-of-sight out-of-mind, confirmation bias, correlation, halo effect, hindsight bias, availability, representativeness, delayed feedback, and confidence, which might be observed from transcripts of geologists discussing information provided by the robot. Compared to highly constrained laboratory studies, a field test allows the results of the study to be directly applied to the problem of understanding what types of error geologists working with a remote robot may make, and what features of the decision-making process are most like to be sources of error. The field test focused on knowledge based decision-making and knowledge based errors, because they tend to be the most frequent error type [12]. Consequently, the objective of this study is to measure the error-rates for these eleven error-associated behavior characteristics during the October field test.

Methods

The field test focused on expert decision-making in robotic geology and was designed to mimic the Mars Exploration Rover (MER) Missions, which launched during the summer of 2003. Three experts in planetary geology participated in the three-day field test. There were two components to the field test, a remote field site with virtual rover and a mission control room. Each daily field test consisted of a six-hour, simulated mission during which the geologists analyzed the field site. The first data set delivered to the geologists each morning in the mission control room was a two-tier 360° panoramic image of the site. Then in two-hour intervals the geologists requested new data on the field site. The amount of data the geologists could request reflected the file sizes for each data type and the total amount of returnable data from the actual MER rovers.

After completing the mission, the geologists traveled to the field site where they identified any differences between what they saw in the control room and what they saw in the environment. In order to reduce biases based on information from the previous day, each new mission was of a different field site.

The field test used two forms of data collection. Video and audio recordings served as passive data collection of the geologists' decision-making behaviors. These two sources provided a transcript of the entire session. Interruptions and interviews every ten minutes with the geologists served as active data collection of their behaviors. During the interviews, the geologists enumerated their current activities, hypotheses, and conclusions.

The combined data sources provided a list of hypotheses and conclusions reached by the geologists during the field test as well as the information they used to support each of the hypotheses and conclusions. The list was then compared to the differences the geologists found when they visited the field site. For this study, any difference between the geologists' interpretations in the control room and the field site was taken to be an error.

Results

The 19,005 lines of audio transcripts contained a total of 424 hypotheses with supporting information. To develop the list of hypotheses, two people independently highlighted sentences in the transcripts and interview notes that contained a conclusion, hypothesis or observation, based on working definitions agreed upon before the analysis began. A comparison of the resulting lists indicated a 98 % correlation between the two analysts. The highlighted sections were reduced to a single list of conclusions, hypotheses and observations, from which specific decision-making behaviors were identified along with errors identified in the field.

For the experiment, a conclusion was categorized as knowledge-based if it involved the use of two or more pieces of supporting information. The geologists reached fifty-one knowledge-based conclusions during the duration of the field test. Each of these were classified by the eleven error-associated behavior characteristics. Each conclusion fit the classification of more than one characteristic (Table 1).

Error-Associated Behavior Characteristics (E-ABC)	Frequency of E-ABC for Each Knowledge-Based Conclusion	Frequency of Knowledge-Based Error for each E-ABC	Error rate	Frequency of All Errors for Each E-ABC
Selectivity	41	2	4.9 %	4
Workspace limitation	10	1	10 %	6
Out-of-Sight, Out-of-mind	1	0	0%	4
Confirmation bias	1	0	0 %	1
Correlation	26	2	7.7 %	3
Halo Effect	0	0	--	0
Hindsight bias	16	2	12.5%	0
Availability	51	3	5.9%	7
Representativeness	0	0	--	0
Delayed feedback	0	0	--	0
Confidence	0	0	--	1
Total	51	3	5.9 %	20

Table 1: *The sub-categories of knowledge based behavior defined by Reason (1990) along with the number of times the behavior was demonstrated by the subjects and the number of times it produce an error.*

At the field site, the geologists identified twenty errors in their interpretation. These included the observation that there was far more basalt than expected, that clear evidence for a streambed in the exposed rock layering had been missed, for example. Of these twenty errors, twelve were possibilities explored but rejected or forgotten by the geologists. The other eight errors were not discussed in the control room (i.e. they were missed by the geologists). Of the twelve errors that had been considered in the transcript, only three fit the classification of a knowledge-based conclusion. The behaviors used in the three cases of error that were knowledge-based decisions are listed in column 3 of Table 1, indicating how often a particular Error-Associated Behavior Characteristic contributed to an error. The last column of Table 1 lists all the times a behavior led to any of the twenty observed errors.

Discussion

The error rates for the different error-associated behavior characteristics suggest that none are strictly indicative of an error. Experts frequently employ these decision-making behaviors to reach correct conclusions. The behaviors are deployed in many situations, but in some of the situations, they lead to undesired results. In many others, the same behaviors help the observer come quickly to the correct decision, making the same characteristics attributes. Why similar situations can produce different results indicates there may be a different explanation for the underlying cause of the errors.

Looking at errors from a different angle, further conclusions can be drawn. A majority of the errors seen involved some aspect of availability. Indeed, availability has been one of the most studied error types (Ross & Sicoly, 1982; Taylor, 1982; Tversky & Kahneman, 1974, 1982). The essence of availability is the correct hypothesis or conclusion never entered into the decision-making process. For example, the geologists didn't consider the possibility that there might be fossilized wood in the environment, so they failed to see it on the first day. On the second and third days, however, they hypothesized that there might be fossilized wood and searched for it. Further research should focus on how to prompt decision makers to include other unavailable hypotheses.

Errors in this experiment were defined as differences between what was seen in the control room and what was seen in the field. The differences were identified by the subjects and therefore, cannot be related to some inability of the subjects. Instead, there must be some information, some visual cue, some impulse present in the field, which was not present in the control room. Identifying what was missing in the control room would produce better, more complete scientific results in future robotic geology missions.

Acknowledgements

The material is based upon work supported by NASA under award No: NAG5-11981. The research described in this (publication or paper) was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Danaher, J. W. (1980). Human Error in ATC Systems Operations. Human Factors, 22(5), 535-545.

Hollnagel, E. (1993). The phenotype of erroneous actions. International Journal of Man-Machine Studies, 39, 1-32.

Lee, K. W., Tillman, F. A., & Higgins, J. J. (1988). A literature survey of the human reliability component in a man-machine system. IEEE Transactions on Reliability, 37(1), 24-34.

Norman, D. A. (1993). Things that make us smart: Defending human attributes in the age of the machine. Reading, Mass.: Addison-Wesley Publishing Co.

Rasmussen, J. (1987). The Definition of Human Error and a Taxonomy for Technical System Design. In J. Rasmussen & K. Duncan & J. Leplat (Eds.), New Technology and Human Error (pp. 23-30). Chichester: John Wiley & Sons.

Reason, J. (1990). Human Error. New York, NY: Cambridge University Press.

Richei, A., Hauptmanns, U., & Unger, H. (2001). The human error rate assessment and optimizing system HEROS -- a new procedure for evaluating and optimizing the man -- machine interface in PSA. Reliability Engineering and System Safety, 72, 153-164.

Ross, M., & Sicoly, F. (1982). Egocentric biases in availability and attribution. In D. Kahneman & P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Sugarman, R. (1979). Nuclear power and the public risk. IEEE spectrum, 16(11), 59-79.

Taylor, S. E. (1982). The availability bias in social perception and interaction. In D. Kahneman & P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. Science, 185(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1982). Availability: A heuristic for judging frequency and probability. In D. Kahneman & P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Woods, D. D., J., J. L., Cook, R. I., & Sarter, N. B. (1994). Behind Human Error; Cognitive Systems, Computers, and Hindsight. Wright-Patterson Air Force Base, Ohio: Crew Systems Ergonomics Information Analysis Center.