Cognitive Models Predicting Surprise in Robot Operators

David Reitter (reitter@psu.edu), **Yang Xu (yang.xu@psu.edu)**
College of Information Sciences and Technology, Penn State, University Park, PA, USA

Patrick Craven (plc35@drexel.edu)
Computing & Informatics, Drexel University, Philadelphia, PA, USA

Anikó Sándor, Chris Garrett, E. Vince Cross, & Jerry L. Franke
Lockheed Martin Lockheed Martin Advanced Technology Laboratories, USA

Keywords: Spatial Path Planning, Robotics, Autonomous Vehicle Operators, Trust

Problem Statement

Robots and other autonomous vehicles have great utility, and even more potential: they go where human drivers can't breathe and where the lives of human pilots would be too costly to risk. Unfortunately, robots and their remote human operators do not always form a cohesive team. When robots make autonomous decisions, operators can be surprised and will, consequently, lose trust in the automation and end up micro-managing the robot. The benefits of partial autonomy are lost in the process.

The project discussed here evaluates ways to communicate robot reasoning to operators when needed (c.f., Kennedy et al., 2007). Its goal is to restore appropriate trust in automation without overloading the operator's attentional resources (c.f., Merritt et al., 2008).

Our approach assumes that misunderstandings between robot and operator are often due to differences in available information about the environment, different decision-making processes, and different levels of experience. Some sensory information may be withheld: the robot might know more than it visualizes, or the human is able to interpret a video feed more accurately than computer vision can. Further, decision-making algorithms and expertise are not synchronized; a robot may have un-inspectable machine-learning models, and an operator might have years of field experience.

A Robot that Explains Itself

Our objective is to enable the robot to convey pertinent information to improve monitoring performance and appropriate trust in the system, via the right modality and at the right time. The experiment discussed asks participants to interact with a system that can explain itself verbally. For instance, it may say *"I see a table and some glass shards on the ground. I planned a path around those obstacles"*. This explanation would allow the operator to accept the reasoning as is, verify it by referencing the simulated video feed, or reject it outright.

The system is designed to preempt operator surprisal by providing explanations at the best moment. It allows the operator to adopt a management-by-exception strategy: monitoring the autonomous vehicles rather than actively controlling them (Franke et al., 2005).

Experiment

The experiment has four conditions: *no explanations*, explanations only by *operator request*, *ongoing detailed* ones, and *selective* ones given when a cognitive model (see next section) detects operator surprisal.

Participants are asked to divide their time between the robot monitoring task (Figure 1) and a secondary, sensory analysis task, which draws away their visual attention. For the primary task, a standard exploration scenario is used with different rooms containing office furniture. It is implemented using a realistic robot simulation and operator interface (Gerkey et al., 2003).

Robots are evaluated by participants using a trust questionnaire (Merritt et al., 2008) and through neglect tolerance and preference ranking. We hypothesize that operators develop more trust in the three explanation conditions, and that they develop trust congruent with robot performance. We expect them to maintain the highest performance at both tasks only in the *selective* (model-driven) condition.

Increased trust is not necessarily a desirable as autonomous systems do make mistakes. In a second experiment, we will concentrate on *appropriate trust.* Here, participants are exposed to a high and a low-performing simulated robot per condition. The low-performing robot makes mistakes in identifying obstacles: it circumnavigates glass shards, while attempting to go through water, while its actual capabilities are the opposite (water only is to be avoided). The ensuing errors have to be corrected by the operator manually.

Application of a Path-Planning Model

How does the cognitive model predict operator surprisal? We have equipped our experiment system with a cognitive model formulated in ACT-R that predicts the operator's cognitive process in planning paths for the robot. When the robot's path deviates from the path that the model predicts, we detect potential for a surprise and issue an explanation. The experiment (see is designed to create situations in which the robot will misunderstand sensory information and plan an inappropriate path.

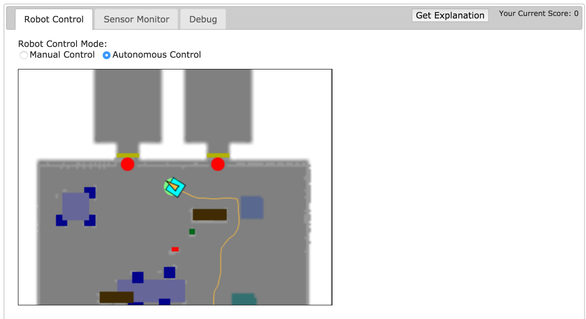

Figure 1: Operator interface, showing pre-defined goals at the top and a path taken by the robot around desks, chairs and some glass debris. In the condition shown, the subject may request a voice explanation.

The cognitive model (Reitter et al., 2010) has originally been developed to fit comparable data: waypoints set by robot operators to carry out an urban search&rescue task. In this setting, robots scout a contaminated office building, circumnavigate walls, cover all rooms, and discover all victims of a fictitious disaster. The insert in Figure 2 shows an itinerary defined by an operator, along with the corresponding model path.

The model predicts the plans an operator would develop for a robot to move from its given location to another given location. As a theoretical rational solution, one may think of a search process that guarantees the shortest workable path. (This standard robotics problem can be addressed via a standard A* algorithm or the more commonly used D*Lite (Koenig et al., 2005).) In contrast, the cognitive model predicts that human operators use a heuristic that selects the straight-line segment available from a given position that reduces the geometric distance to the goal; the initial choice is made at the starting position, and then the algorithm is applied recursively until the destination is reached or backtracking becomes necessary (for models of spatial navigation, compare Fum et al., 2000; Zhao et al., 2013).

The model explains scalability of the task with size of the environment as well as with cognitive load, such as when paths are to be planned for multiple robots (see Figure 2). It was evaluated with automatically generated mazes and on a dataset gained from robot operators that controlled 4, 8 or 12 simulated urban search&rescue robots at a time (Lewis et al., 2007).

Conclusions

As valuable as explanations may be, they can have a downside: cognitive overload and distraction. Therefore, our goal is to provide information when we believe it is necessary during the monitoring task. The experiment is designed to evaluate this approach.

The poster will present our analysis of the empirical results with 40 participants (the experiment has not been

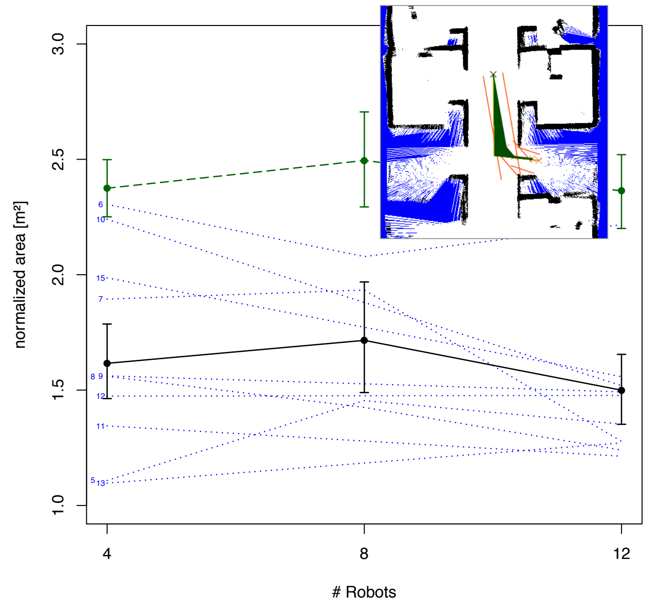

Figure 2: Average model error (solid), model vs. individual subject error (dotted), and avg. baseline model error (dashed) at different operator workload conditions. Insert: Example operator path for a robot (red crosses) and model's prediction for that path; difference area in solid green (from: Reitter et al., 2010).

concluded at the time of writing).

With our approach, we do not design a cognitive model to fit new experimental results. Instead, we use a model that has been evaluated before as a means to predict the expectations of human operators in a realistic task relevant to national defense, safety and security. The experiment helps us analyze explanations as a means to affect trust in autonomous system. It also allows us to evaluate an ACT-R model in an extrinsic setting.

Acknowledgements

This work was funded by the Air Force Research Laboratory. We would like to thank Joe Lyons for his comments.

References

Franke, Jerry L et al. (2005). "Inverting the operator/vehicle ratio: Approaches to next generation UAV command and control". In: *Proc. of AUVSI Unmanned Systems North America* 2005.
Fum, D. and F. Del Missier (2000). "Climbing the mazes: A cognitive model of spatial planning". In: *In Proc. of the Third Int'l Conf on Cognitive Modeling*. Universal Press.
Gerkey, Brian, Richard T Vaughan, and Andrew Howard (2003). "The player/stage project: Tools for multi-robot and distributed sensor systems". In: *Proc. 11th Int'l Conference on Advanced Robotics*. Vol. 1, pp. 317–323.
Kennedy, William G et al. (2007). "Spatial representation and reasoning for human-robot collaboration". In: *AAAI*. Vol. 7, pp. 1554–1559.
Koenig, Sven and Maxim Likhachev (2005). "Fast replanning for navigation in unknown terrain". In: *Robotics, IEEE Transactions on* 21.3, pp. 354–363.
Lewis, Michael, Jijun Wang, and Stephen Hughes (2007). "USARSim: Simulation for the study of human-robot interaction". In: *Journal of Cognitive Engineering and Decision Making* 1.1, pp. 98–120.
Merritt, Stephanie M and Daniel R Ilgen (2008). "Not all trust is created equal: Dispositional and history-based trust in human-automation interactions". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50.2, pp. 194–210.
Reitter, David and Christian Lebiere (2010). "A Cognitive Model of Spatial Path Planning". In: *Computational and Mathematical Organization Theory* 16.3, pp. 220–245.
Zhao, Changkun, Jonathan H Morgan, and Frank E Ritter (2013). "Understanding human high-level spatial memory: An ACT-R model to integrate multi-level spatial cues and strategies". In: *Biologically Inspired Cognitive Architectures* 3, pp. 1–5.