How to Trust Robots Further than We Can Throw Them

David Bruemmer
Idaho National Laboratory
bruedj@inel.gov

Douglas Few
Idaho National Laboratory
fewda@inel.gov

Michael Goodrich
Brigham Young University
mike@cs.byu.edu

Donald Norman
Nielsen Norman Group
don@jnd.org

Nilanjan Sarkar
Vanderbilt University
nilanjan.sarkar@vanderbilt.edu

Jean Scholtz
National Institute of Standards and Technology
jean.scholtz@nist.gov

Bill Smart
Washington University in St. Louis
wds@cse.wustl.edu

Mark L. Swinson
US Army Research Office
mark.swinson@us.army.mil

Holly Yanco
University of Massachusetts, Lowell
holly@cs.uml.edu

ABSTRACT

Panelists with backgrounds in diverse aspects of human-robot interaction will discuss the challenges of human-robot interaction in terms of operator trust. The panel will showcase experiments and case studies that highlight the importance of operator trust. Each panelist provides a unique perspective on the role of trust in mobile robot applications and offers unique insight on how we can help build trust for a future generation of mobile robots. Panelists have been drawn from defense, entertainment, industry, transportation and energy sectors. The panel will discuss cases where humans were too willing to place trust in robot systems and others where humans have been unwilling or unable to trust robot behavior. In each instance, panelists will point to current shortcomings (i.e., interfaces, communications, robot intelligence) and plans to address these limitations in the future.

Author Keywords

Human-Robot Interaction, Trust, Robot Control

ACM Classification Keywords

B.1.3 Control Structure Reliability, Testing, and Fault-Tolerance

PANEL OVERVIEW

Humans have high expectations for mobile, intelligent robots. We have been told by film, fiction and television that they will infiltrate our homes and places of work, changing the landscape of human experience. Where are these robots and what should we do if we see one? This panel will investigate issues of trust surrounding implementation and use of mobile robots in a variety of different applications. The panel will showcase experiments and case studies that highlight the importance of operator trust. Panelists will discuss how operator trust has affected the use and efficacy of their robots and will show videos of their robots in action to illustrate ways in which robots can fail. Panelists will also explain where issues of human trust of robots arise and the underlying causes of this distrust. Each panelist will have an opportunity to describe what they are doing and what they think should be done in general to support greater trust between humans and robots.

Some of the robots described represent a new frontier of research where robots and humans work together as peers. This teaming between human and robot can be synergistic or, at times, counterproductive, depending on the level of human trust and their familiarity with the system. Conversely, creation of system trust may depend on the ability of the robot to act predictably and support different levels and frequencies of operator intervention.

Human-robot interaction is a complex and fascinating area for discussion. It leads to questions of a philosophical and ethical nature, and the panel will encourage this tendency. For instance, some of the robots discussed have the ability to say "No!" to human operators in situations where the safety of the robot or environment is at stake. Some panelists will argue that when configured appropriately for the task and when well-understood by the user, the ability to refuse dangerous or futile commands can be a most helpful asset to operators. Others will argue that robots should always be subservient to human demands.

Teams of humans and robots present a new kind of automation that questions whether the human should retain leadership over all aspects of the task. The paradigm of "human supervision" may change, but different approaches have different foci. Should human supervision give way to interaction that reflects true team interaction, or is full automation that writes the human out of the loop sufficient? Other approaches that reflect some other middle ground may be more appropriate. The panel will particularly interest those working in areas of artificial intelligence, interface design, usability testing and control theory.

SUMMARY OF PANELISTS' VIEWS

Each panelist provides a unique perspective on why humans do and do not trust robotic systems and offers his or her own tactics for how to build greater trust.

Dr. Smart has developed a robotic wedding photographer and is investigating ways to provide the human with a window into the capabilities and limitations of the robot. In order to provide this window, he believes in exploiting established cues such as making eye-contact or saying "Say Cheese!" in order to influence human expectations and build trust in the system.

Dr. Sarkar believes that it is not enough for the human to understand the robot. He is interested in how robots can be given a means to understand the human. Using a wide variety of bio-feedback sensors, he provides the robot with a window into the human operator's mental state including their level of attention and frustration. The robot can use this information to adapt itself to the needs of the user and build a greater level of trust.

Dr. Norman believes that in order to build trust, we need to humanize robots, providing them with meaningful emotions and personality. His research has demonstrated that human error is not typically a function of just the human, but rather a function of the system in question not accurately or effectively facilitating the user's understanding of how the system works. One method to do this is to provide to the human a paradigm for interaction that the user already knows: human-human interaction, which would require providing robots with meaningful personality traits. This philosophy can also give insights to how humans and robots should interact, namely through "natural" communication channels such as speech. Dr. Norman also has shown that training does not always overcome trust issues arising from poor design.

Dr. Goodrich works with remote tasks for mobile robots and is also involved with driver assist tools for use within the automotive industry. He will illustrate how too much trust can prompt the user to ignore the system, resulting in poor performance. He will also give examples of how too little trust can lead to over monitoring and an inability for the user to accomplish secondary tasks. Dr. Goodrich's research has focused on tools to improve the situation awareness of the human when interacting with robots. Dr. Goodrich points to training as a means for building trust and believes that trust can be negotiated throughout a task.

Dr. Swinson will discuss trust within the context of bomb disposal robots, where the technical problems of teleoperated manipulation and the lack of trust in autonomy has prompted the development of robots that can scale the level of autonomy to support different levels of user trust.

Dr. Scholtz will discuss the dynamics of trust for different interaction roles of humans with robots, including supervisor, operator, mechanic, teammate, and bystander. Each role requires a different level of trust and imposes very different needs in terms of information and understanding. Dr. Scholtz is interested in understanding how the presentation of information can be made appropriate for each role. Her research has often focused on how a true understanding of the task should influence creation of tools for that task.

Mr. Few believes that trust will ultimately be based on reliability and that it is the system architect's responsibility to provide the foundation of trust by creating dependable or at the very least predictable systems. In contrast to other panelists who focus on modes and methods of interaction, Mr. Few believes that trust depends largely on the reliability of the robot subsystems and the fault-tolerance of the robot architecture that unifies these components.

In contrast to many of the panelists' work with remote operations, Dr. Yanco has developed a semi-autonomous robotic wheelchair – a system that demands complete user trust. These lessons are highly applicable to the domain of human-robot interaction. For robots that work in the same space as humans, it is necessary that the system not only be safe, but that it actually enact behaviors in such a way as to *seem* safe. In order for users and observers to feel confident, we must develop ways for the system to provide status, explain its behavior and ask for help when needed.

Mr. Bruemmer is part of a research group working to create robot architectures that support human-robot teaming in dangerous environments Usability testing of mobile robots at the INEEL with his colleagues, Dr. Julie Marble and Mr. Douglas Few, has shown that although system performance can be improved by sharing control between the human and robot, the possibility for counterproductive interaction is also high. User distrust may lead to a 'fight' for control. The focus of this research has been that understanding robot actions and intentions may be more important than roboto reliability or performance.

How to Trust Robots Further than We Can Throw Them: A Panel Discussion

Contact: David Bruemmer (bruedj@inel.gov / 208 526 4078)

Researchers are all too ready to tell you what their robots do well. Perhaps that is one of the reasons so many people are waiting with baited breath for the promises of science fiction to come on the market. We think it's far more telling to hear (and watch) how robots fail. To kick off the panel discussions, those panelists who work closely with robots will provide a video clip or narrative illustration of their own robot's limitations. This "robot follies" segment will serve to communicate current robot limitations and underscore the difficulty of establishing trust between humans and robots. Every panelist agrees that trust is a crucial issue, affecting the broad range of projects represented from helpful robots such as automated wheelchairs and robot wedding photographers to homeland defense operations including rescue robots and bomb disposal robots. Panelists responsible for some of the best examples of human-robot interaction in the world today will give the inside scoop on what they themselves do and do not trust their robots to do. Panelists who are not roboticists will explain what they would *never* trust a robot to do and why. Each panelist will use an example of undeserved trust and an example of when deserved trust was withheld to illustrate how trust affects the use of robotic systems. Of course, each panelist will have an opportunity to describe what they are doing or what can be done in general to build greater trust.

After showing clips of their robots and providing a couple of background slides that give the audience a perspective on their work, panelists will be interviewed with a number of questions appropriate for their area of expertise, such as:

- What areas of your work (or which robot applications) are most affected by issues of trust?
- Give examples of how trust has affected the utility or deployment of your (a) robotic system.
- What is the biggest hurdle that inhibits trust of robots in general?
- What steps have you taken to increase trust?
- What far-reaching steps can be taken to create systems that people can and will trust?
- Do you think there are or should be inherent limits to the trust that people place in robots?

After each panelist has been interviewed by the moderator, the audience will have an opportunity to ask questions about their work with robots or about their take on issues of human-robot trust.

In particular, the panel will investigate issues surrounding implementation and use of robot autonomy. Some of the robots described represent a new frontier of research in the area of intelligent robots. These robots have the ability to take initiative to accomplish high-level task objectives such as "Search this room for human victims!" or "Take me to the library!" On the other hand, users often want to step in during the task and micro-manage the behavior of the robot. This team effort can be synergistic or, at times, counterproductive, depending on the attitude of the human user and their familiarity with the system. Conversely, success depends on the ability of the robot to act predictably and support different levels and frequencies of operator intervention. Most importantly, success depends on the ability of the human and robot to recognize and respond to each other's limitations.

Human-Robot interaction is a complex and fascinating area for discussion. It will invariably lead to questions of a philosophical and ethical nature and the panel will encourage this tendency. For instance, some of the robots discussed have the ability to say "No!" to human operators in situations where the safety of the robot or environment is at stake. Some panelists will argue that when configured appropriately for the task and when well-understood by the user, the ability to refuse dangerous or futile commands can be a most helpful asset to operators, especially when the robot informs the user clearly and in a timely fashion. Others will argue that robots should always be subservient to human demands.

True teams of humans and robots present a new kind of automation where the notion that the human retains leadership over all aspects of the task is now in question. The paradigm of "human supervision" may well give way to a promising, but simultaneously frightening future of "peer-to-peer interaction" where authority is given to whichever team member is most appropriate, be they human or machine. On the other hand, the panel is also focused on simple robots for everyday use. Several people on the panel point out that we will always have a need for robots that serve as passive tools. Some panelists will argue that we certainly do not want our lawn mower to 'talk back.' Others will argue that maybe we do. After all, our lawnmower may know more than we do about mowing lawns. Without question, these are issues, which will be of interest across the entire human factors community. In particular, this panel will be of interest to those who work in areas of automation, interface design, usability testing and control theory. The panel will tie in very well with the Special Area focus on Robotics & Transport, which is actually chaired by one of the panelists.

Each panelist provides a unique perspective on why humans do and do not trust robotic systems and offers unique insight on how we can help build greater trust.

David Bruemmer (bruedj@inel.gov / 208 526 4078) is the organizer of this panel and will serve as the interviewer / moderator for the panel discussion. At the Idaho National Engineering and Environmental Laboratory, he leads research and development efforts for a variety of application domains including landmine detection; remote characterization of high radiation environments; spill finding and perimeter formation; urban search and rescue and military operations. Previous work includes technical consulting for the Defense Advanced Research Projects Agency in the areas of human-robot interaction, humanoid robots and swarming robots.

Panelist Summary: Sharing control is not always a good idea. Human-subject experiments at the INEEL have found that although system performance can be improved by sharing control between the human and robot, the possibility for counterproductive interaction is also high. User distrust may lead to a 'fight' for control. On the other hand, many users readily relinquish control even when the robot is performing poorly and needs help. A good control system should not encourage unwarranted trust, but rather should build understanding by exposing limitations as well as strengths. Interestingly, understanding may be more important than trust. Users will find ways to utilize a faulty robotic system they understand, but will often refuse to use a high-performing, reliable robot that they do not understand. In order to build understanding, there's nothing like hands on experience. One way to build trust with operators is have them try to drive 200-300lb robots into a wall at high speeds. Trust is built as the robot repeatedly halts two inches from the wall. Another method of instilling trust is to put the robot in a remote environment and then kill the video feed, forcing the human to rely on the robot's capabilities – a trial by fire.

Bill Smart (wds@cse.wustl.edu / 314 935-4749) is an Assistant Professor of Computer Science and Engineering at Washington University in St. Louis. He is the co-director of the Media and Machines Laboratory, which conducts research in the areas of Mobile Robotics, Computer Vision, Machine Learning, and Computer Graphics. His research interests cover the fields of mobile robotics and machine learning, and include long-term autonomy and reinforcement learning for robot control.

Panelist Summary: In the context of mobile robotics, trust is most affected by the robot's ability to live up to the (often implied) contract of behavior with the human observer. Dr. Smart has developed and deployed a robot photographer system at a number of real-world events. This system wanders about social gatherings, taking candid, well-framed shots of people at the event. Although the robot has been very successful, one of the greatest challenges is inflated expectations of the human observer. For example, the perceived "eye-contact" made as the robot's pan/tilt cameras track a human face leads many people to infer that the robot is interested in them or what they are doing. Instead of treating it as a passport photograph booth on wheels, they treat it more like a human. People shout or wave at it and become confused and frustrated when the robot fails to do what they tell it. One way to address this issue is to find out what a typical human's expectations are, and program the robot to conform to them. For example, people expect photographers to say "say cheese" before taking a picture, and for a flash on the camera to go off when the picture is taken. Dr. Smart has found that adding this functionality, even if it serves no functional purpose, raises people's understanding and consequently their trust in the system. The second approach is to influence the model of behavior used by human observers. Dr. Smart is currently designing studies that will test how the robot can influence human expectations.

Donald A. Norman (don@jnd.org / 847 498-4292) is Prof. Emeritus of Cognitive Science at the University of California, San Diego, former Vice President of the Advanced Technology Group at Apple Computer, and an executive at Hewlett Packard. He is co-founder of the Nielsen Norman group, a consulting firm that promotes the development of human-centered information appliances. He is the author of 13 books including "The Design of Everyday Things," "Things That Make Us Smart," and "The Invisible Computer," which Business Week calls "the bible of post-PC thinking."

Panelist Summary: Most robotics efforts aim for high performance. For Dr. Norman performance is not enough. He believes that if we are to trust robots and integrate them into our lives, we must first humanize robotics technology, creating a human-centered family of intelligent robots that we can easily interact with and learn to use. Instead of merely simulating emotion or trying to exploit simple cues, Dr. Norman believes that robots should be able to respond to body motion, to the speed and forcefulness of our activities, to our thoughts and feelings, to where, how and when we look, to speech, sound, music and touch. Interaction with intelligent robots should be enjoyable. Dr. Norman believes that the emphasis should be on cooperative, equal participation of humans and machines. In such a system, trust becomes a very complex issue. Dr. Norman believes that emotion can play an important role in mediating trust between humans and robots. Emotion helps to define and prioritize goals, and also provides a convenient means to communicate internal state. In fact, in contrast to others on the panel, Don Norman believes that intelligent behavior cannot take place without emotion.

Nilanjan Sarkar (nilanjan.sarkar@vanderbilt.edu / 615 343-7219) is an assistant professor in Mechanical Engineering, and Computer Engineering at Vanderbilt University where his research interests include robotics, human-robot collaboration, control theory, and rehabilitation engineering. Dr. Sarkar is currently an associate editor of the IEEE Transactions on Robotics and Automation and was a guest editor for IEEE/ASME Transactions on Mechatronics.

Panelist Summary: As in human teams, trust must be based in an ability to understand and monitor the behavior of the other team member. In contrast to other panelists who are most concerned with the ability of the human to monitor and understand the robot, Dr. Sarkar has chosen to focus on the ability for the robot to monitor the human. Consider a remote tele-robotic task where the robot could vary its level of trust based on the operator's level of attention or change its behavior when an operator became frustrated or respond to an operator who had been injured. Dr. Sarkar's research has developed ways to infer the affective states of humans by using wearable biofeedback sensors to measure physiological data such heart rate, skin conductance, etc.. This data is analyzed using real-time signal processing and machine learning techniques to infer the underlying psychological state of the person, which is then used to affect a change in robot strategy and behavior. Such mental state inference, if performed reliably, will provide a means to build trust by adapting to the needs and limitations of the human.

Jean Scholtz (jean.scholtz@nist.gov / 301 975-2520) is a computer scientist at the National Institute of Standards and Technology. Dr. Scholtz is a member of ACM, SIGCHI, IEEE, and UPA. She has been active in the CHI conference for many years and has served on the SIGCHI executive board. She is the special area chair for Robotics and Transport for CHI 2004. Her research interests are in the evaluation of complex, interactive systems. As a program manager at the Defense Advanced Research Projects Agency, Dr. Scholtz started a human-robot interaction program, Synergistic Cyber Forces. She is currently studying human-robot interactions related to on-road driving robots, and robotic search and rescue activities.

Panelist Summary: Dr. Scholtz thinks we cannot simply talk about operator trust in general. People have different expectations depending on the different interaction roles they play: supervisor, operator, mechanic, teammate, and bystander. For each of these roles, different information is required for effective human-robot interaction. Dr. Scholtz's research is focused on identifying the types of information and interactions needed to support each role. An unresolved question is the way that the presentation of this information affects the user's perception and trust. Raw sensor data such as ladar and sonar is difficult for human individuals to interpret. This brings up some interesting issues including whether users trust human-like sensory information over more unfamiliar modalities? Will data fusion and abstraction increase or decrease user trust. What happens when sensory information from several sources conflict? What length of time must individuals work with a given robot to establish trust? During the panel Dr. Scholtz will use observational data and video clips from a number of search and rescue competitions to illustrate issues where trust plays a key role in effective human-robot interaction.

Mark L. Swinson (mark.swinson@us.army.mil / 919 549-4204) is the Director for Research and Technology Integration at the US Army Research Office where he is responsible for overseeing a $300 million/year research and technology integration portfolio. Prior to this assignment he was the deputy director of the Intelligent Systems and Robotics Center at Sandia National Laboratories. Before coming to Sandia Dr. Swinson served as the DARPA program manager for several robotics efforts before serving an appointment as deputy director of DARPA's Information Technology Office. A retired U.S. Army Colonel, Dr. Swinson was the Army's senior military roboticist and served as the program manager of the Joint Tactical Unmanned Aerial Vehicle (JTUAV) program, and was the founding technical director of the Joint Project Office for Unmanned Ground Vehicles. He is also a recognized expert in the area of humanoid robotics and intelligent systems, and is a member of the IEEE Intelligent Systems magazine's editorial board.

Panelist Summary: As a retired Army Colonel, Dr. Swinson can talk about applications where robot behavior can directly affect human life. One of the most challenging areas for trust is in the area of high consequence manipulation. A canonical example is bomb disposal. While bomb disposal robots have been around for some time, their utility has been limited by the robot's inability to effectively function in teleoperation, while being insufficiently trustworthy to handle tasks autonomously. Dr. Swinson will discuss how an ability to dynamically shift the level of robot autonomy from teleoperation to autonomous behavior execution, can dramatically reduce the obligatory level of operator training and experience, while offering the operator an opportunity to gradually increase confidence in autonomous operations. Within this paradigm, the human maintains control, but can assign responsibilities to the robot as necessary.

Holly Yanco (holly@cs.uml.edu / 978 934-3642) is an Assistant Professor in the Computer Science Department at the University of Massachusetts Lowell. She earned her MS and Ph.D. degrees from MIT in 1994 and 2000. Her research interests include adjustable autonomy, assistive technology, multiple robot teams, and human-robot interaction. She is a co-editor of the book "Lecture Notes in Artificial Intelligence: Assistive Technology and Artificial Intelligence." Dr. Yanco was co-chair of the AAAI Robot Competition and Exhibition (2001 and 2002) and was the chair of the special interest group on Assistive Robotics and Mechatronics of the Rehabilitation Engineering Society (2001-2003).

Panelist Summary: In contrast to many of the panelists' work with remote operations, Dr. Yanko has developed a semi-autonomous robotic wheelchair – a system which demands complete user trust. The current system takes directional commands from its user, and executes them using obstacle avoidance and path following algorithms. Future systems could involve much more autonomy, where a user sitting in his office would trust the wheelchair to drive him to the copy room or to a library across the street. In addition to the user, the medical personnel, caregivers and family members must also trust the system. Especially at first, the user's stress level will increase in situations that seem dangerous, even if they are not. For example, in one user test, the system came close to the edge of a curb. Although the wheelchair did turn in time, the test subject was very nervous and almost halted the chair. For robots that work in the same space as humans, it may be necessary that the system not only be safe, but that it actually enact behaviors in such a way as to *seem* safe. In order for users and observers to feel confident, we must develop ways for the system to provide status, explain its behavior and ask for help when needed. However, we cannot impose additional burdens especially since users may be non-verbal, have sight problems or decreased mental capacity, or may not have fine motor control.

Michael A. Goodrich (mike@cs.byu.edu / 801 422-6468) received the Ph.D., M.S., and B.S. degree in electrical and computer engineering from Brigham Young University in 1996, 1995 and 1992, respectively. He has been a research associate with Nissan Research and

Development, Inc., Cambridge. Since 1998, he has been with the Computer Science Department at Brigham Young University where he teaches and leads research in human-robot interaction, multi-agent coordination, modeling and controlling intelligent systems, decision theory, and machine learning.

Panelist Summary: Unlike other panelists who talk about trust as a subjective quality, Dr. Goodrich has attempted to measure trust experimentally. If you want to know how much users really trust the robot, measure how much time they will devote to a secondary task. For semi-autonomous robots, a lack of trust often prompts operators to over-monitor and devote little time to a secondary task. The opposite problem can be even worse. Under-monitoring occurs when operators place too much trust in the robot. For example, in an experiment designed to illustrate the benefits of sonar-assisted shared control, many users felt uncomfortable allowing the system to drive on its own and often stepped in to take control. The solution was to explicitly instruct the subjects to "point the robot in the direction that you want it to go and *trust* that it will do what you want." This simple instruction produced not only higher task performance but also greater neglect tolerance. Unfortunately, this simple instruction may not be appropriate for higher levels of automation (e.g. "Point the robot in the direction it will go and trust that it will use its munitions correctly"). In order to support user trust, Dr. Goodrich suggests developing dialogue-based systems wherein trust is negotiated; using training to build better system understanding and using interfaces that support trust-based interactions.

Douglas A Few (fewda@inel.gov / (208 526 3077) is currently a Senior Research Scientist for the Idaho National Engineering and Environmental Laboratory. He recently won 1st place in the 2003 Robot Rescue Competition and has a lot to say about the need for trust in such remote operations. Before working at INEEL, Mr. Few worked at iRobot, which develops a variety of robots for human use including autonomous vacuums, research robots, military robots, robot dolls, etc. He is interested in integrating robotic systems, robot software architectures, and novel interaction tools to accomplish critical tasks.

Panelist Summary: Mr. Few believes that it is the system architect's responsibility to provide the foundation of trust by creating dependable or at the very least predictable systems. In contrast to other panelists who focus on modes and methods of interaction, Mr. Few believes that trust depends largely on the reliability of the robot subsystems and the fault-tolerance of the robot architecture that unifies these components. In terms of hardware, robotic systems include a host of sensors, communication systems, computing platforms, electronics and mechanical components all of which can fail in their own unique ways. Sometimes roboticists choose to say that these sticky problems are not theirs, and simply assume accurate and robust sensing, communication, positioning, and mechanical operation. Most of today's robotic systems employ complex algorithms for mapping, localization, obstacle avoidance and other behaviors which are built on the premise of continuous valid sensor data and no component failures. Mr. Few assures us that robot systems will fail. He believes that the effect of these failures on user trust is directly dependent on the system's ability to handle the failures at a system architecture level.