Building and Maintaining Trust Between Humans and Guidance Robots in an Emergency

Paul Robinette[1,2], Alan R. Wagner[2], and Ayanna M. Howard[1]

[1]School of Electrical and Computer Engineering
Georgia Institute of Technology
[2]Georgia Tech Research Institute
Atlanta, GA, USA 30332
probinette3@gatech.edu, Alan.Wagner@gtri.gatech.edu, ayanna.howard@ece.gatech.edu

Abstract

Emergency evacuations are dangerous situations for both evacuees and first responders. The use of automation in the form of guidance robots can reduce the danger to humans by both aiding evacuees and assisting first responders. This presents an interesting opportunity to explore the trust dynamic between frightened evacuees and automated robot guides. We present our work so far on designing robots to immediately generate trust as well as our initial concept of an algorithm for maintaining trust through interaction.

Introduction

When alarms sound, strobes flash, and smoke fills the air there is no time to consider the trust model between yourself and the person directing you to the nearest exit. Evacuees assume that firefighters, police officers, and uniformed store employees are good sources of knowledge in an emergency evacuation regardless of any personal experience with these forms of authority. Unfortunately, first responders require time to respond to the scene and employees are not always willing to put their own lives in danger to search for and aid evacuees inside a burning building. Some fires are so extreme that firefighters are not even able to enter the building by the time they arrive on the scene (Grosshandler et al. 2005).

As robots begin to enter commercial and residential use it becomes possible to imagine a solution using automated guides to solve problems caused by panic and uncertainty in emergencies. A robot otherwise tasked with floor cleaning or food service can be retasked in an emergency to provide guidance to frightened or panicked evacuees. How will evacuees respond to these robots? How can we use results from experiments in this domain to create more general models of trust between humans and automated agents?

In preliminary experiments we have begun to answer the following two questions:

- How many people need to trust a robot as a guide to significantly increase survivability in an evacuation? (Robinette and Howard 2011b; Robinette, Vela, and Howard 2012)
- How many people tend to trust a robot as a guide in an evacuation? (Robinette and Howard 2012)

For the purposes of these experiments we defined our trust metric as the probability that the human will follow a robot to an exit. This prior research utilized the robot's appearance and behaviors to instill trust in notional evacuees. Here we have developed a refined system that allows evacuees to build trust in the robot by making the robot behave and appear as an authority figure. The robot then maintains trust by talking to the evacuee.

Related Work

Robots are becoming popular tools for traditional search and rescue missions. Using feedback from their studies, Bethel and Murphy suggested using voice communication to reassure victims and music when there is no information to communicate (Bethel and Murphy 2008).

In previous evacuation robot research, robots with directional audio beacons were deployed in optimal positions to reach as many people as possible (Shell and Mataric 2004; Shell and Mataric 2005). These robots were shown to decrease the total amount of time to evacuate in a simulation of an emergency.

Several studies have been performed investigating how people react in emergency situations. Sime found that individuals with strong ties to a group were less likely to panic and try to escape in a selfish way (Sime 1983). Another study determined that individuals chose the main entrance as their preferred exit in a simulated emergency (Benthorn and Frantzich 1999).

Defining Trust

Wagner defines trust as "a belief, held by the trustor, that the trustee will act in a manner that mitigates the trustor's risk in a situation in which the trustee has put its outcomes at risk" (Wagner 2009). He also denotes four conditions for trust:

1. The trustee does not act before the trustor.
2. The outcome received by the trustor depends on the actions of the trustee if and only if the trustor selects the trusting action.
3. The trustor's outcome must not depend on the action of the trustee when selecting the untrusting action.
4. The value of fulfilled trust is greater than the value of not trusting at all, is greater than the value of having one's trust broken.

This definition of trust is particularly appealing for emergency situations because it directly deals with risk. Robots in an emergency situation are not attempting to reward humans for their compliance with directions; they are trying to mitigate risk to human life.

Wagner's main example is the trust fall: the trustor puts himself at risk by falling backwards. The trustee chooses whether or not to catch him. The evacuation scenario is analogous to this example: the evacuee is presented with a situation involving great risk (e.g. a fire emergency) and must decide if she should place her fate in the robot (Figure 1). The evacuee therefore chooses whether or not to follow the robot. The actions available to the robot, on the other hand, are to safely guide a person to the exit or to not safely guide the person to an exit. These actions are meant to convey the consequences of the robot's decisions, not the robot's intentions. Hence, if the robot becomes lost while attempting to follow a safe path to the exit then, from the evacuee's perspective, the robot has not selected the correct actions leading to a safe exit.

With respect to Wagner's conditions for trust, the evacuee must choose whether or not to follow the robot before they can know with certainty whether or not the robot safely reaches the exit. Thus, condition one is satisfied.

Condition two relates to the person's dependence on the robot. It notes that the evacuee's risk depends on the actions and ability of the robot, if the evacuee chooses to follow the robot. This condition is also satisfied.

Condition three notes that if the person chooses not to follow the robot, then the outcome he or she receives is largely independent of the robot. In this case if the evacuee chooses not to follow the robot then the robot's actions and ability do not impact the person's risk. This condition is also satisfied.

The fourth condition notes that it is better (in some sense) for the evacuee if she chooses to follow the robot and it successfully leads her to an exit, than to have not followed the robot at all, than to have followed the robot and not been led to an exit. The advantage of following a robot to an exit versus finding the exit by oneself may not be apparent. During disasters people tend to have a large cognitive and emotional load. Following the robot may reduce this load and allow a person to simply follow an object to a safe exit rather than have to manage a potentially large number of navigation decisions (i.e. is this stairwell safe, which direction is the closest exit, etc.). Hence, we argue that this final condition is also satisfied.

In addition to adopting Wagner's definition of trust, we have also adopted his use of outcome matrices to represent interactions requiring trust (Figure 1). An outcome matrix lists pairs of actions by two agents along with their corresponding rewards for each agent. In the robot-guided evacuation domain we restrict the actions of the trustee (the evacuee) to either following robot guidance or not. Likewise, we restrict the actions of the trustor (the robot) to either providing good guidance (being a good trustor) or providing poor guidance. For the purposes of this work we assume that the robot will not intentionally provide poor guidance, but we do offer a scenario where the evacuee might think the robot would attempt to deceive him or her.

An outcome matrix for an ideal evacuee in a robot-guided evacuation can be seen in Figure 1. This outcome matrix was generated by considering the emergency from the evacuee's perspective. The two actions available to the evacuee are to follow the robot or ignore the robot. The two actions available to the robot are to guide the evacuee safely or not.

The values for the outcome matrix are idealized estimates for the scenarios discussed in the text. The robot's outcomes are listed as an 'X' to indicate that these values do not impact our analysis. Negative values are meant to indicate a cost to the individual. Positive values indicate a reward. In Figure 1, the value of -5 for the Ignore action represents the cost of searching for the exit without the robot's help. The 0 value for the (Follow, Safe Guide) action pair depicts the fact that being safely lead out of danger by the robot does change the person's state. The value of -20 for the action pair (Follow, Unsafe Guide) is meant to illustrate the cost of the harm done to the person by remaining in the disaster area.

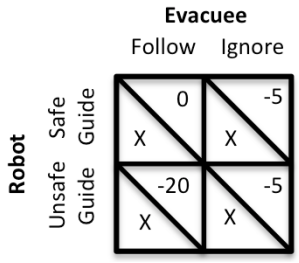

Figure 1: General Evacuee's Outcome Matrix

Building Trust

Bickman found that requests made by someone wearing an official uniform are more likely to be followed than those made from someone wearing a commercial uniform or civilian clothes (Bickman 1974). This research even showed that people would tend to respond negatively to such requests in a hypothetical scenario, but follow the request in an actual interaction. Our existing designs for evacuation robots take advantage of this uniform effect by appearing to be an official piece of fire equipment. An outward appearance similar to a uniform or recognizable piece of fire equipment increases an evacuee's confidence in the robot, therefore increasing trust in the robot. We developed two candidate designs for user simulation: (Robinette and Howard 2011a)

- Robot 1: red and white striped robot with typical North American exit signs to provide directional information and a fire department seal to provide authority (Figure 2a)
- Robot 2: white cylindrical robot with clear red arrows to provide directional information and the words "Emergency Evacuation Robot" along the back in red as a statement of purpose (Figure 2b).

In trials published in (Robinette and Howard 2012), users reported that they immediately understood the purpose of these robots and generally understood that they were supposed to proceed in the direction of the robot's arrows. If the user can quickly understand the intention of the robot then the user will deem selecting the trusting action as less risky and be more likely to choose that action.

(a) Robot 1 **(b) Robot 2**

Figure 2: Original Robot Designs

New robot designs have been developed to test some concerns among users that Robot 1 did not look like a serious emergency responder and that Robot 2 was easily lost in smoke. Both robots have been given a red exterior, with the text color on Robot 2 changed to white. This both increases the visibility of each robot and reinforces the perception that the robot is a certified emergency responder.

Additionally, both robots have been given arms. The arms are articulated such that they can perform standard gestures used by emergency responders to guide individuals. These arms will provide further clarity of direction for evacuees by using familiar dynamic gestures to supplement the static guidance symbols already on the robot. It is also believed that the arms will make the robot appear more active and slightly more human. This, in turn, should reduce the evacuee's assessment of the risk associated with following the robot.

In response to some concerns that the robots appeared jerky and indecisive, (Robinette and Howard 2012) we are following the advice of other rescue robot researchers and designing the robots to move smoothly along a constant trajectory (Bethel and Murphy 2008).

Maintaining Trust

As an initial test, we created a simulated 3D evacuation disaster in software (Robinette and Howard 2012). We recruited fifteen volunteers to each perform seven evacuation scenarios both with and without robotic guidance. Our results indicated that all evacuees will follow guidance robots in the first scenario presented; however, this simulation was low stress and simple, so real evacuees may react differently. In real disasters many people will risk personal injury and even death to find members of their group (Sime 1983). In our tests we have also received feedback that some evacuees felt the robot behaved in an untrustworthy manner.

To account for these situations, we have developed a model that informs the robot of possible motivations for an evacuee or a group of evacuees who disregard guidance suggestions. First, the robot attempts to attract trust in its guidance from an evacuee using methods listed in the previous section. Second, the robot determines if the evacuee appears to be following its guidance. Next, the robot asks a series of questions to the evacuee to find out why the evacuee will not follow. The robot also observes the evacuee to gain additional information. Using this information, the robot determines the current mental mode that best describes the evacuee. Finally, the robot attempts

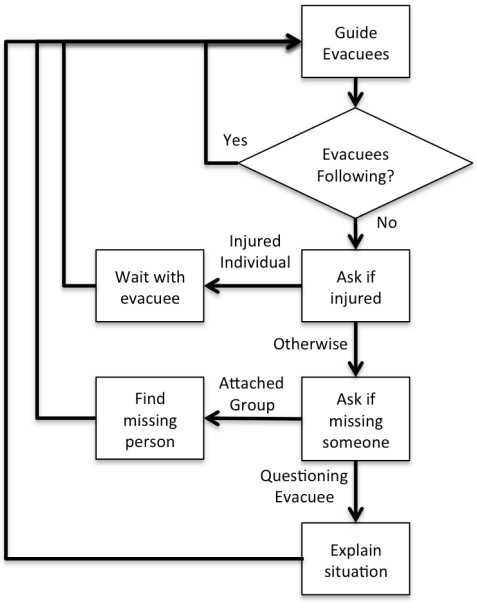

Figure 3: Example Decision Tree for Maintaining Trust

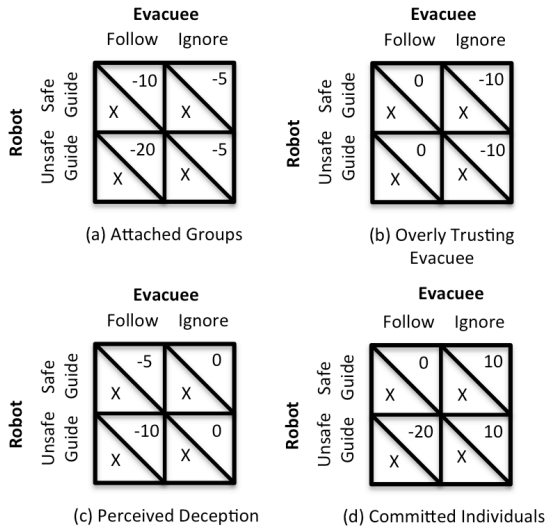

Figure 4: Outcome Matrices illustrating the decision problem faced by different evacuees. Matrix (a) depicts the increased cost of following the robot and leaving an attached group behind. Matrix (b) illustrates the evacuee's belief that following the robot does not entail risk. Matrix (c) depicts the added cost of following a robot that might be acting deceptively. Matrix (d) depicts the evacuee's confidence in their own evacuation plan as reward for ignoring the robot.

to raise the evacuee's confidence in the robot's direction so that it can return to its guidance mode. An example of this decision process can be seen in Figure 3.

By acting predictably and explaining its behavior the robot will allow evacuees to form a simple model of the robot's decision processes. If the robot can maintain consistency with this model then evacuees will have increased confidence in the actions of the robot. The model also considers the evacuee's perception of their own risk in the situation. The robot gives certain assurances, such as informing the evacuees of the estimated time to exit and giving status updates on the arrival of emergency personnel, in order to reduce this perceived risk. Increasing confidence in the robot and decreasing the perceived risk of the situation will allow evacuees to increase their trust in the robot.

Several cases where evacuees may not follow guide robots are shown below. A selection of their outcome matrices is shown in Figure 6.

Attached Groups

Interviews with victims after disasters indicate that many evacuees will wait in a burning building at great personal risk to ensure that their family and close friends are together and safe (Sime 1983). In this case, the evacuee may perceive the robot to be trustworthy but will still ignore the robot's guidance until the rest of his or her group can be found. The matrix for an attached group has a large cost associated with following the robot. This cost is a reflection of the negative consequences of leaving people behind. In some cases this cost makes the follow command a less attractive option than remaining at the site. The outcome matrix for this case can be seen in Figure 4a. Whether or not the evacuee believes that the robot will lead them to safety may become irrelevant because of the cost associated with leaving others behind.

The robot starts to identify this case by noticing that an evacuee is not following and is not injured. The robot then asks the evacuee if he or she is looking for members of his or her group. If this is the case, then the robot now knows that the evacuee does not intend to leave the building until all members of the group are found.

The robot can take several actions at this point. The first action is to help the evacuee look for the rest of the group. It is unknown if an evacuation robot will actually be able to help parents locate children in an emergency, but the model accounts for this possibility. If the group member is found then the evacuee should be willing to evacuate and he or she will have greater trust in the robot due to the robot's assistance.

The second action is that the robot can communicate with other robots and first responders to determine if the group member has already evacuated. If this is the case then the robot can transmit video or other proof to the evacuee, which should convince the evacuee to follow the robot to safety.

The final action is to attempt to convince the evacuee that the risk is too great to continue to remain in the building. This is unlikely to work on subjects such as a

parent looking for a lost child, but it is the only action remaining for the robot to take.

Injured Individuals

Some evacuees may be injured such that they are not able to evacuate. The robot can identify this case by noticing injuries and by asking the immobile evacuees if they are injured. Ideally, the robot's perception will be capable of determining obvious injuries on a human, but regardless of perception, showing concern for the evacuee's well being will increase the personal nature of care. This case has no relevant outcome matrix because we assume that the injured individual is unable to take any action.

The most obvious action that the robot can take in this situation is to move the evacuee to safety. There are numerous practical issues with this action as it both requires the robot to have the ability to lift an adult human and requires the robot to be able to identify how to lift the evacuee without causing further injury. Robots will only perform this action when there is immediate danger to the victim or when an emergency responder can supervise.

The simplest action that a robot can perform in this case is to wait with the evacuee until first responders can help. This allows the robot to act as a communication device between the injured evacuee and first responders so that first responders can effectively prioritize rescue operations. As suggested in (Bethel and Murphy 2008), the robot will be able to play music for the injured victim.

Questioning Robot-Guided Route

Many evacuees in our simulated evacuation reported that they felt the robot chose an exit further away than necessary (Robinette and Howard 2012). Some even noticed that the robot would pass nearby exits and signs in favor of more distant exits. This was used as an experimental condition to determine how trusting evacuees would be, but in a real scenario such questions require a response from the robot. Existing research has shown that evacuees tend to prefer the main entrance when exiting (Benthorn and Frantzich 1999), sometimes even to the point of crowding and trampling at that door (Grosshandler et al. 2005). Robot guides need to be able to convince evacuees to move towards less obvious exits. The definition for trust presented in the third section informs us that if the robot indicates that it has an incentive to get the person to the exit, then the evacuee will view the action of following the robot as less risky and be more likely to do so.

Here, once again, the process begins when the robot recognizes that the person is not following. The robot then attempts to determine why. The robot listens for questions and complaints about the chosen route. The robot may also watch the evacuee to see if he or she is confused or if he or she attempts to turn towards more familiar exits.

The only action that the robot can take in this case is to explain its actions. The robot will choose the closest exit that is safe, so it can explain its decision process to the evacuee as necessary. Explaining its actions will allow the evacuee to learn about the robot's decision-making process and thus begin to trust it more. If the evacuee still refuses to follow the robot then the robot will contact first responders in the hopes that they can convince the evacuee to take a better exit.

Overly Trusting

In our previous simulation (Robinette and Howard 2012), one evacuee reported that one of the robots seemed faster and was thus more trustworthy. These individuals view the risk of following the robot as less than it may actually be. The possibility that evacuees may trust the robot too much, is a relevant aspect for consideration and hence, should be addressed. The outcome matrix is shown in Figure 4b.

Generally, fast movements by the robot would be considered desirable. Still there are some situations in which the robot may need to be slowed down for practical reasons. For example, the robot may slow down to prevent congestion at exit points. This would only be used when it is safe, such as when the building is being evacuated as a precaution due to other events.

We expect that slow robots will be viewed as less trustworthy. If the robot recognizes that evacuees are moving faster than itself (a likely scenario) then the robot can attempt to provide the evacuee with enough instruction that he or she can proceed without further guidance.

When it is necessary for the evacuee to stay with the robot then the robot will explain the situation and reasons to slow down. The robot will demonstrate that it has more knowledge of the situation than the evacuee by explaining the crowded conditions at the exit. The robot will also explain the exact nature of the emergency so that the evacuee understands there is sufficient time to take a slower pace. By being honest and explaining its actions the robot will be able to increase trust from the evacuee.

Perceived Deception

One test subject in our simulation reported that he felt the robot was trying to deceive him by leading him away from the exit (Robinette and Howard 2012). This individual thus believed that the risk of following the robot is greater than the risk of not following the robot. We have no intention of developing a deceptive robot guide, but it is still a valid concern for evacuees. Note that the evacuee believes that the robot's motivation is exactly reversed, thus the robot does not intends to safely guide him to an exit. This outcome matrix is shown in Figure 4c.

If the robot notices that the evacuee is reluctant to follow, then the robot can inquire as to why. If the evacuee mentions deception by the robot then the only action the robot can take in this case is to continue to describe the situation to the evacuee. Some individuals cannot be convinced that the robot intends to help and is capable in its task.

Committed Individuals

In previous work, we performed experiments by varying the number of individuals who truly believed they knew the best path to the exit, regardless of actual knowledge (Robinette, Vela, and Howard 2012). These individuals exist in real evacuations as well and the robot must be prepared to convince them of the error in their judgment. The outcome matrix for a committed individual resembles an extreme case of an individual questioning robot guidance (Figure 4d).

The robot can identify committed individuals because they will ignore any attempts by the robot to guide them. The robot will respond to this by approaching the individuals and explaining why their chosen path is undesirable. It is assumed that even the most committed individuals will change their mind when faced with a sufficiently dangerous situation.

Conclusion

In an emergency evacuation, individuals need to instantly decide who to trust and who to ignore. We have presented revisions to our design for emergency evacuation robots that enable them to immediately attract trust from evacuees. Using feedback from an initial experiment and relevant literature we have also presented a concept for an algorithm that can identify evacuees who are not responding to robot guidance, categorize them into the correct evacuee case, and recommend actions for increasing trust in the robot (Robinette and Howard 2012).

Our algorithm for maintaining trust is applicable to many situations where one agent wishes to increase trust from another agent. The emergency evacuation domain is a particularly compelling application for this algorithm because of the great risk but low reward for the evacuee.

Our algorithm and associated trust model are still in their early stages of development. Future work includes testing this system in our evacuation simulator by intentionally breaking an evacuee's trust and then evaluating how well the system rebuilds trust.

References

Benthorn, L. and Frantzich, H. 1999. Fire alarm in a public building: How do people evaluate information and choose an evacuation exit? *Fire and Materials*, vol. 23, no. 1, pp. 311–315.

Bethel, C. L. and Murphy, R. R. 2008. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, And Cybernetics Part C*, 38(1):83–92.

Bickman, L. 1974. The social power of a uniform. *Journal of Applied Social Psychology* 4(1):47–61.

Grosshandler, W.; Bryner, N.; Madrzykowski, D.; and Kuntz, K. 2005. Report of the technical investigation of The Station Nightclub Fire. Technical report, National Institute of Standards and Technology.

Helbing, D.; Johansson, A.; and Al-Abideen H. Z. 2007. Dynamics of crowd disasters: An empirical study. *Physical Review E*, vol. 75, no. 4, p. 046109.

Robinette, P., and Howard, A. 2011a. Emergency evacuation robot design. In *ANS EPRRSD - 13th Robotics & Remote Systems for Hazardous Environments and 11th Emergency Preparedness & Response.*

Robinette, P., and Howard, A. 2011b. Incorporating a model of human panic behavior for robotic-based emergency evacuation. In *RO-MAN, 2011 IEEE*, 47–52. IEEE.

Robinette, P., and Howard, A. 2012. Trust in emergency evacuation robots. In *10th IEEE International Symposium on Safety Security and Rescue Robotics (SSRR 2012)*.

Robinette, P.; Vela, P.; and Howard, A. 2012. Information propagation applied to robot-assisted evacuation. In *2012 IEEE International Conference on Robotics and Automation*.

Shell, D. and Mataric, M. 2004. Directional audio beacon deployment: An assistive multi-robot application, in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 3, pp. 2588–2594, IEEE.

Shell, D. A.; and Mataric, M. J. 2005. Insights toward robot-assisted evacuation. *Advanced Robotics*, 19(8):797–818.

Sime, J. D. 1983. Affiliate behaviour during escape to building exits. *Journal of Environmental Psychology*, 3:21–4.

Wagner, A. R., and Arkin, R. C. 2011. Recognizing situations that demand trust. *RO-MAN, 2011 IEEE*.