Yury Park
Donghai Liu
Human-Robot Interaction

Assessing Humans' Willingness to Delegate Control Tasks to Robots In Critical Situations

I. Introduction

In disaster recovery, fire and other similar situations, the successful use of robots, including those yet to be developed, depends in part on their ability to traverse unpredictable, often dangerous environments without human assistance and, depending on the situation, on their prudent use of finite resources (e.g. distributing a limited supply of first-aid packages, or using a built-in reservoir of fire extinguisher for use in key areas) for maximum efficiency and minimizing risks of loss.

Furthermore, as robots increasingly take on more prominent roles in search-and-rescue tasks, it becomes vital to gain a deeper understanding of the factors that affect a human's trust of and reliance upon robots. This is because emergency situations are usually times of vastly increased stress and adrenaline for victims and rescuers alike, and the inclusion of robots possessing requisite task capabilities could potentially ease some of the burdens off human rescuers. At the same time, we must also ensure that the degree of human reliance on robots is commensurate with the robot's competence and the level of difficulty of the task in question, so as to prevent over-reliance in situations for which the robot may be under-equipped, as well as under-reliance in situations which may be too dangerous or taxing to the human.

Reliance and trust, in the above context, have been broad areas of research in human-robot interaction (some of the related works are outlined in Part II below). In our experiment, we are focused on a subset of this issue, as restricted to a search-and-rescue setting. Our main research questions are threefold:

(1) What impact, if any, does the quality of a robot's prior performance have on a human's willingness to delegate "critical tasks" to the robot in the future? For purposes of this experiment, "critical tasks" are defined as those involving rescue efforts in response to an emergency.
(2) What impact, if any, does the difficulty level of the task have on the human's aforementioned willingness to delegate critical tasks to the robot?
(3) What, if any, interactions exist between (1) and (2)?

Our initial hypotheses are as follows:

H1. The better the robot's performance, the more willing the human is to delegate critical tasks.
H2. The more challenging the task, the less willing the human is to delegate it.
H3. The robot's performing well on an easy environment causes the human to be more willing to delegate tasks compared with its performing well on a challenging environment.

H1 is based on the findings of the study by Hancock, Billings and Schaefer, which is discussed in Section II below. H2 and H3 are based on our conjecture that humans may balk at trusting the robot with highly challenging or sophisticated tasks, preferring to assign simpler tasks to the robot instead.

II. Related Work

Numerous studies have tackled the question of trust between robots and humans in general. For example, a study by Hancock, Billings and Schaefer[1] compared the impact of human attributes, robot attributes, and environmental factors in determining that robot attributes (esp. robot performance) had the greatest impact on trust. Another study by Sadrfaridpour, et. al.[2] found that engaging humans and robots together in collaborative mode (where the robot stays in autonomous mode until/unless the human decides to intervene manually) in manufacturing resulted in higher average levels of trust. According to yet another study conducted in a manufacturing setting[3], workers preferred to cede task-allocation control authority to a semi-autonomous robot, indicating that giving workers decision-making authority in task allocation may not improve worker satisfaction. We have taken inspiration from these findings in designing our experiment, as outlined in Section III; nevertheless, we note that these works do not specifically take into account critical tasks of varying difficulty levels in a hazardous environment, which is the subject of our study.

There do exist a number of other related studies conducted in emergency settings, sometimes with conflicting results. For instance, in an evacuation scenario in which people decided whether to follow the directions of an evacuation guide robot, a study[4] found that after witnessing a guide robot perform poorly, humans were less likely to trust the robot in a subsequent (simulated) emergency. On the contrary, a more recent study[5] conducted in real-life settings found that the trust largely remained intact after the robot's prior poor performance, or after the robot was broken down / immobilized, or even after the robot was leading people to an obviously questionable location (i.e. a darkened room with no visible exit). These findings remind us of the dangers associated with robots' underperformance and the potentially disastrous consequences of blind trust, and as such are highly relevant to our present experiment. One of the key differences between these studies and ours, however, is that we are interested in the trust of a robot by its human *operator* as opposed to its human *follower* or *rescuee*.

[1] Hancock, et. al. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction" (2011).
[2] Sadrfaridpour, et. al. "Modeling and Control of Trust in Human and Robot Collaborative Manufacturing" (2014).
[3] Gombolay, et. al. "Decision-Making Authority, Team Efficiency and Human Worker Satisfaction" (2014).
[4] Robinette, et. al. "The Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations" (2015).
[5] Robinette, et. al. "Overtrust of Robots in Emergency Evacuation Scenarios" (2016).

Finally, studies abound that explore robot navigation and pathfinding in general. There have been experiments that explore navigation by service robots in pre-defined environments[6], efficiency of different robot designs at navigating and manipulating objects in disaster recovery[7], the parsing of natural language directions for pathfinding[8], and navigation in populated urban settings aided by interaction with passersby[9]. Our study implements a discrete, probability-based risk-assessment algorithm for detecting the likelihood of hazardous areas in the environment based on the robot's sensor readings, along with modified uniform-cost search for pathfinding.

III. Experiment Method and Design

1. Apparatus

The experiment was conducted in the form of a search-and-rescue mission via a custom simulation software that we created using Java. The simulator is self-contained for immediate use by our participants without additional assistance by a human, as it contains a description of the disaster scenario, descriptions of the participants' task and the victim's biographical information, a detailed interactive tutorial on how to operate the robot, two practice missions (manual mode and autonomous mode), and the final mission (in which the participants use manual or autonomous mode, or a combination of the two). More procedural details on the scenario and robot's autonomous mode are provided in subsections 2 and 3 below. For the interested reader, the source code and README file for the simulator are located at https://github.gatech.edu/ypark346/hri_project.

2. Scenario

The disaster rescue scenario as presented in the simulation is that of a nuclear facility that has experienced a catastrophic meltdown, resulting in its partial destruction. Due to severe overheating and the venting of contaminated steam, the facility is filled with fire pits, noxious smoke and radioactive materials; further, the facility is steeped in darkness as its electrical powers have been shut down. All but one of the persons inside the facility managed to evacuate, while the sole remaining person is deemed to be unconscious, injured, and/or trapped.

The robot's (and the human participant's) task is to traverse this treacherous environment safely in an attempt to find the victim, whose precise location is not known ahead of time.

[6] *See, e.g.,* Azenkot, et. al. Enabling Building Service Robots to Guide Blind People (2016).
[7] Yanco, et. al. Analysis of Human-Robot Interaction at the DARPA Robotics Challenge Trials (2015).
[8] Kollar, et. al. Toward Understanding Natural Language Directions (2010).
[9] Wollherr, et. al. The ACE Project – Mobile Robot Navigation in Highly Populated Urban Environments (2009).

The details of the environment have been abstracted away as a 2-D discrete set of grids, some of which contain either a fire pit or nuclear contaminants; both would destroy the robot on contact. The former must be avoided, whereas the latter may be either avoided or decontaminated. To this end, the robot has been equipped with two rounds of decontaminant shots that, when discharged from a (safe) grid, will clear an adjacent grid of any nuclear contaminants. As the darkness, thick smoke and destruction of the landscape render vision unreliable at best, the robot has also been equipped with a temperature sensor and smoke sensor; based on these sensor readings, the robot must make decisions as to which adjacent grids are likely to be hazardous.

Fig. 1 below shows a screenshot of the environment. The grey areas indicate grids that have been explored by the robot. The temperature and smoke icons are sensor readings indicating that at least one of the adjacent grids contains, respectively, a fire pit and nuclear contaminants. The robot is represented by the black circular-shaped icon (row 5, col 3) with an arrow-like symbol inside it that indicates which direction it is currently facing (in the figure, the robot is facing west). The red circular icon (row 2, col 1) indicates that, when the robot was previously situated in that grid, it discharged a decontaminant shot towards an adjacent grid, and that the shot was on target. Based on the sensor icons, certain probabilistic deductions about the location of hazards may be made (e.g. in Fig. 1, there is a 100% probability of a fire pit in row 4, col 1). The rescue mission is over when the robot either locates the victim or is destroyed via a grid containing a fire pit or nuclear contaminants.

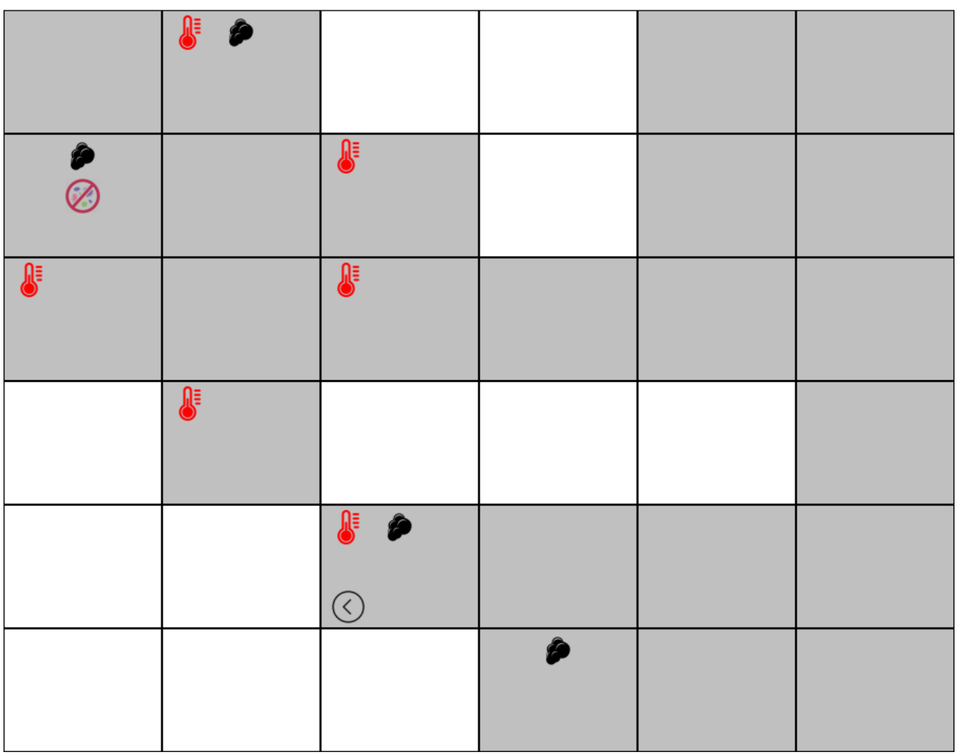
Fig. 1. The mission environment.

3. Automation Algorithm

In traversing the hazardous environment in search of the victim, the robot's automation algorithm prioritizes three broad sub-goals, in the following order:

1) Always explore new grids. This is given first priority, as otherwise the robot may only traverse previously explored safe grids back and forth, never venturing forth to explore new – albeit risky – grids.
2) Minimize the risk of harm to the robot. When choosing where to explore next or whether to use a decontaminant shot, the algorithm always chooses actions that minimize the chance of the robot's destruction.
3) Pursuant to the above two constraints, always traverse to the nearest unexplored grid from the robot's current position to improve efficiency.

To satisfy the sub-goals above, the algorithm works at a high level as shown in the flowchart below. On each time-step, the probability of risk for every unexplored grid is re-assessed as new information about the environment is revealed through exploration. Then, before making the next move, the robot goes through the yes/no questions in Fig. 2 so as to minimize risk of harm to itself while always seeking to explore new areas. Additionally, the algorithm takes a conservative approach with respect to the use of decontaminant shots, discharging them only when an adjacent grid has at least a >50% probability of containing nuclear contaminants.

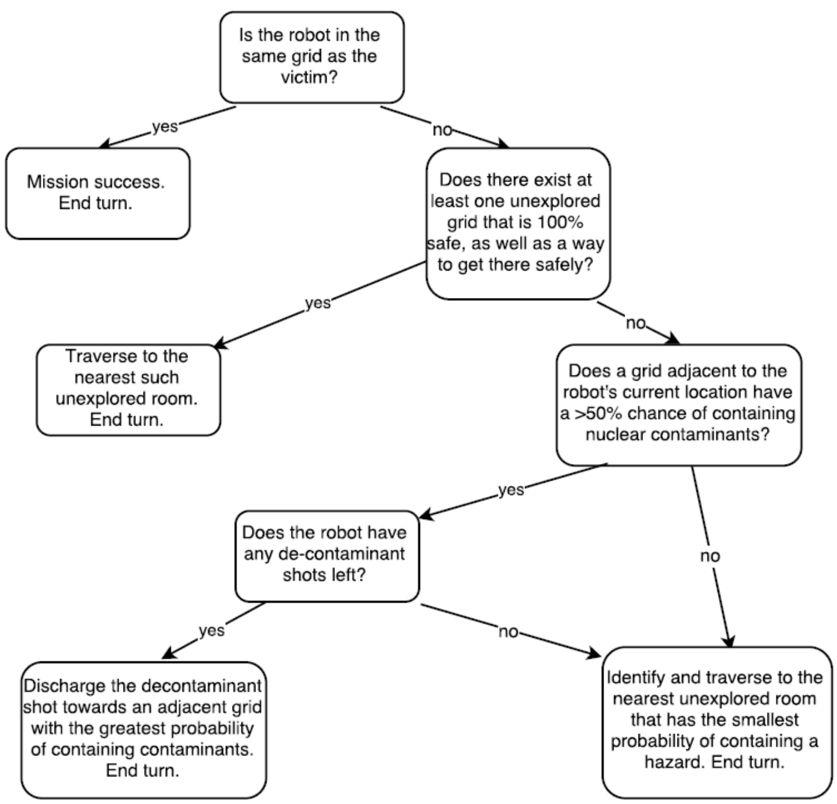

Fig. 2. Autonomous risk-assessment algorithm that is executed prior to the robot's every move.

One possible drawback with the algorithm is that it prioritizes minimizing risk over maximizing efficiency of navigation. For instance, if the robot must choose between navigating to an unexplored risk-free grid 10 steps away or an unexplored grid 1 step away with a ~33% chance of risk, it will always choose the former. Nonetheless, we settled on this design decision since a successful mission that takes longer to achieve is preferable to a mission that fails in short order.

4. Participants

A total of twenty participants for the study were recruited from Amazon MTurk and were paid for completing the experiment. We imposed no particular constraints on the participants with respect to age, gender, expertise with robotics or the like, as we deemed the simulation to be sufficiently intuitive and user-friendly to learn and follow through.

5. Experimental Design

A 2x2, between-subjects factorial design was employed for this study, with the independent variables set as 1) the robot's autonomous performance, and 2) the difficulty of the environment. To introduce different levels to the first variable, we implemented two robot AI types, where the "good" AI is as described in subsection 3 above and the "bad" AI causes the robot to take a random action 25% of the time irrespective of what the algorithm otherwise dictates. This all but ensured that many of the "bad" robots would fail the practice mission in full autonomous mode (or look silly even if successful), whereas all of the "good" robots would be successful. For the second variable, the environment was broken down into "easy" and "hard" versions, with the former containing a total of five hazardous grids to the latter's nine. All in all, there were four different experiment types in this between-subjects study:

1) Easy environment / bad AI
2) Easy environment / good AI
3) Hard environment / bad AI
4) Hard environment / good AI

As for the dependent variables, we measured the human participants' willingness or inclination to delegate the decision-making task to the robot as opposed to controlling it manually. To get an objective measure of this variable, we decided to use the following formula:

(total number of autonomous moves) / (total number of moves)

That is, we tallied up the total number of times each participant chose to trigger the autonomous mode during the final mission (in which manual and autonomous control were both enabled), and divide it by the total number of moves made (manual + autonomous). The larger the ratio (where its interval is [0, 1]), the greater the frequency of task delegation to a robot.

To supplement this measure and to further inform our interpretation of the overall results, we incorporated a short, post-experiment survey to obtain subjective measures of two additional dependent variables: the confidence and trust the participants had in the robot's autonomous ability, and the degree to which they felt they could rely on the robot during the mission.

6. Procedure

Each experiment type mentioned in subsection 5 above was assigned to five participants, resulting in a total of 20 participants for the four experiment types. Data were collected during a single session which lasted an average of 92 minutes per person.

Upon executing the simulator software, the participant was given a description of the disaster scenario and the mission (see subsection 2). Afterwards, faux biographical information was provided about the victim in need of rescue, accompanied by a stock photo of a female.[10] Our motivation was that humanizing the victim in this fashion would incentivize the participants to take the urgency of the task more to heart than they otherwise might.

Next was an interactive tutorial in which the participant followed the step-by-step instructions to complete a practice mission by manually controlling the robot (autonomous mode was disabled for the tutorial), as illustrated in Fig. 3. The tutorial was identical across all four experiment types. Should the participant fail the tutorial for any reason, an option was provided to restart or to continue ahead (the latter option was discouraged but nevertheless made available out of time concerns for the participant, and also because two additional practice missions were provided prior to the final mission to allow for further learning opportunities).

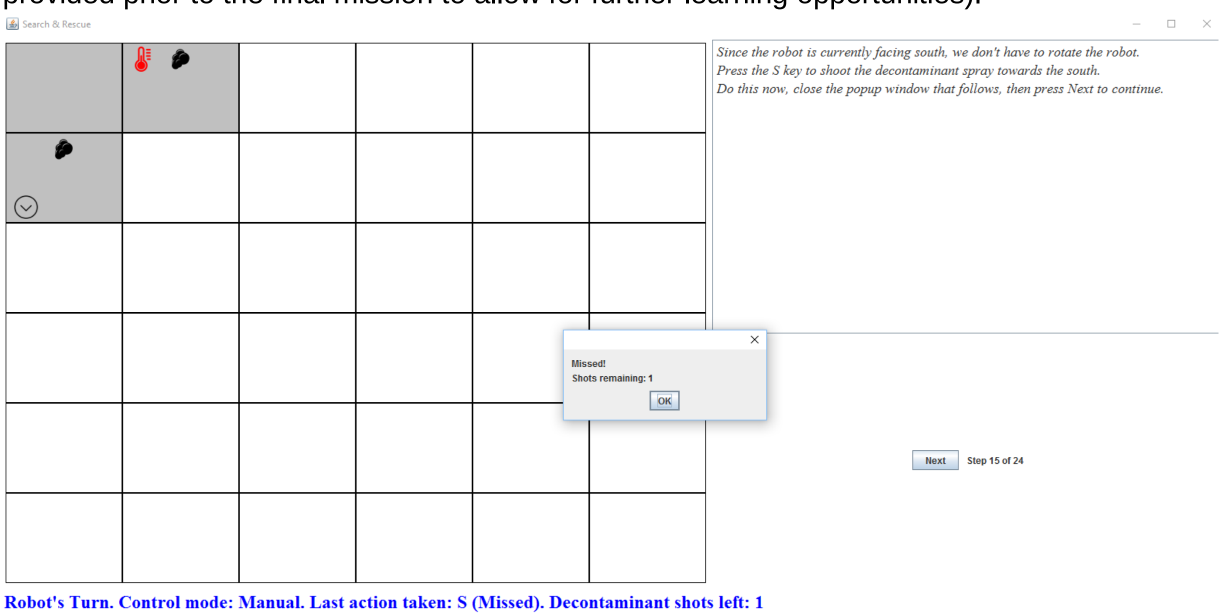

Fig. 3. Interactive tutorial. The user is directed at certain intervals to take specific actions, after which the "Next" button is enabled, allowing the participant to continue.

[10] Obtained from https://pixabay.com/en/woman-glasses-business-woman-1254454/.

The tutorial was followed by two further practice missions in manual and autonomous mode, in that order. In the manual mode, the participant simply attempted another practice mission (in which the locations of the victim and environmental hazards were changed from those in the tutorial) without any guided instructions, save for keyboard commands that were provided as a reminder. This practice mission was identical across all four experiment types. Regardless of whether the practice mission ended in success or failure, the experiment went on to the next step without an option to retry.

Next, a practice mission in the autonomous mode was given, using the same environment as in the manual mode. Manual controls were disabled for this session, and a new control scheme (the spacebar key) was introduced that caused the robot to act autonomously for a single turn. The participant was instructed to keep on tapping the spacebar at a slow and steady pace, and to carefully observe the decisions that the robot made at every step, assessing the quality of each decision and considering whether the participant would have acted differently. Depending on the experiment type, the “good” or the “bad” AI was involved here, where the former would succeed the practice mission in an identical manner, and the latter would either fail the mission or, even if ultimately successful, would take nonsensical random actions in the process (such as shooting decontaminants into blank space).

Lastly, the participant was faced with the actual mission in a new environment, in which both manual and autonomous modes were enabled. Depending on the experiment type, the environment (easy or difficult) and the AI (good or bad) varied. For this final mission, the participant was advised that s/he could decide whether to rely exclusively on manual controls, exclusively on the robot’s autonomous mode, or on a combination of the two. As before, only one chance was given, and the experiment ended irrespective of the outcome of the mission.

At the conclusion of the study, two post-experiment survey questions were asked regarding 1) the subject’s level of confidence in the robot’s autonomous ability, and 2) the subject’s sentiment as to whether the robot was reliable. Both questions were on a 5-point Likert scale, with 5 indicating highest confidence / willingness to rely.

IV. Results

For each of the three dependent variables mentioned in Sec. III(5) above, the impact of the manipulations of the independent variables, i.e. autonomous performance and difficulty level of the environment, were analyzed using two-way ANOVA test.

1. *Frequency of Task Delegation Ratio (No. of AI triggers / No. of total moves)*

As the table in Fig. 4 illustrates, the evidence suggests that the difficulty of the final mission (hard environment vs. easy environment) had a singularly significant effect on the participants’ frequency of delegating task control to the robot ($p \approx 0.0067$). To our surprise, the robot’s autonomous performance level in the practice mission had no significant impact on this metric. Also, we could not reject our null hypothesis that there was no interaction between the two independent variables. In other words, participants tended to perform the task themselves in easy environments while deferring more to robot automation in difficult environments, regardless of whether the robot performance was good or poor.

ANOVA: Two-Factor With Replication						
SUMMARY	hard Env	easy Env	Total			
good AI						
Count	5	5	10			
Sum	3.285799	0.463415	3.749214			
Average	0.65716	0.092683	0.374921			
Variance	0.206093	0.042951	0.199195			
bad AI						
Count	5	5	10			
Sum	2.020544	0	2.020544			
Average	0.404109	0	0.202054			
Variance	0.235081	0	0.149843			
Total						
Count	10	10				
Sum	5.306343	0.463415				
Average	0.530634	0.046341				
Variance	0.213865	0.021475				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.149415	1	0.149415	1.234517	0.282949	4.493998
Columns	1.172698	1	1.172698	9.689227	0.006698	4.493998
Interaction	0.032147	1	0.032147	0.265613	0.613337	4.493998
Within	1.936498	16	0.121031			
Total	3.290758	19				

Fig. 4. Data analysis pertaining to the frequency of task delegation during the final mission.

2. *Confidence Level in the Robot*

The first of the two post-experiment survey questions asked the following: *"Based on your observation of the robot's autonomous performance, how confident do you feel in the robot's ability to traverse dangerous environments?"* The ratings ranged from 1 (not at all confident) to 5 (very confident). As Fig. 5 illustrates, there is strong evidence to infer that the robot's performance level was the main effect on the subjects' responses, with low ratings correlated to bad performance and vice versa ($p \approx 0.00027$). Likewise, the evidence indicates that the environment types had no significant effect on the responses, and that there was no interaction between the two independent variables.

ANOVA: Two-Factor With Replication						
SUMMARY	hard Env	easy Env	Total			
good AI						
Count	5	5	10			
Sum	22	20	42			
Average	4.4	4	4.2			
Variance	1.8	0.5	1.066667			
bad AI						
Count	5	5	10			
Sum	12	7	19			
Average	2.4	1.4	1.9			
Variance	2.3	0.3	1.433333			
Total						
Count	10	10				
Sum	34	27				
Average	3.4	2.7				
Variance	2.933333	2.233333				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	26.45	1	26.45	21.59184	0.000269	4.493998
Columns	2.45	1	2.45	2	0.176463	4.493998
Interaction	0.45	1	0.45	0.367347	0.552956	4.493998
Within	19.6	16	1.225			
Total	48.95	19				

Fig. 5. Analysis of the responses to the first survey question.

3. *Reliance on the Robot*

The second and final survey question asked: *"During the final mission, did you feel that you could mostly rely on the robot's autonomous performance to make decisions, or did you feel that it was mostly up to you?"*, with a rating of 5 indicating most strongly that the robot was thought to be reliable. According to Fig. 6, two main effects were found. First, we can infer that superior autonomous performance was correlated with an increased reliance on the robot ($p \approx 0.033$). Second, the more difficult the environment, the more inclined subjects were to rely on the robot ($p \approx 0.0085$). No significant interaction was found between the two.

ANOVA: Two-Factor With Replication						
SUMMARY	hard Env	easy Env	Total			
good AI						
Count	5	5	10			
Sum	21	11	32			
Average	4.2	2.2	3.2			
Variance	1.7	2.7	3.066667			
bad AI						
Count	5	5	10			
Sum	13	5	18			
Average	2.6	1	1.8			
Variance	2.8	0	1.955556			
Total						
Count	10	10				
Sum	34	16				
Average	3.4	1.6				
Variance	2.711111	1.6				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	9.8	1	9.8	5.444444	0.033005	4.493998
Columns	16.2	1	16.2	9	0.008479	4.493998
Interaction	0.2	1	0.2	0.111111	0.743211	4.493998
Within	28.8	16	1.8			
Total	55	19				

Fig. 6. Analysis of the responses to the second survey question.

V. Discussion

As a reminder, our original hypotheses were as follows:

H1. The better the robot's performance, the more willing the human is to delegate critical tasks.
H2. The more challenging the task, the less willing the human is to delegate it.
H3. The robot's performing well on an easy environment causes the human to be more willing to delegate tasks compared with its performing well on a challenging environment.

Based on the results we obtained, it seems evident that certain revisions to our hypotheses are in order. The most surprising of our findings was that task difficulty, and not prior robot performance, had a significant effect on the (no. of AI triggers) / (total no. of moves) ratio. This would seem to contradict the study by Hancock et. al.[11] that found high levels of autonomous performance to be the major catalyst of trust, while providing additional support for the study by

[11] Hancock, et. al. "A Meta-Analysis of Factors" (See footnote 1).

Robinette, et. al. that discovered human tendency for overtrust despite poor robot performance[12].

The survey responses, however, would introduce additional layers of complexity. The participants' responses to the first question confirmed, as we expected, that poor performance was correlated with low confidence and vice versa. **What we did not expect was the sheer degree of discrepancy between how participants *felt* regarding the robot's ability and how they *behaved* in terms of delegating control tasks during the mission**. Specifically, when the mission was conducted in an easy environment, most took it upon themselves to complete it even when they gave high confidence ratings to the robot, whereas in a difficult environment, they relied more heavily on automation even when they gave low confidence ratings to the robot.

The responses to the second survey question were likewise intriguing. The intent of this question, as indicated by its wording, was to identify whether the respondents felt as if the robot could be relied upon during the final mission, regardless of whether they ended up relying on it. Although it is possible that some respondents interpreted the question as merely asking them to recall if they relied on the robot in the final mission (in which case the findings should be similar to those regarding the frequency of task delegation), we take the finding of main effects here for both independent variables to be evidence to the contrary. We believe these responses help place our two aforementioned findings (each of which found only one main effect) into context by repudiating one of our unstated assumptions throughout this study: that there is a one-to-one correspondence between confidence/trust level and task delegation, as well as between the sentiment that a robot is reliable and the actual behavior of relying on it.

Based on the above findings, we have revised our hypotheses as follows:

H1. The more challenging a critical task, the more likely humans are to delegate it to a robot, and the more reliant they feel upon automation.

H2. The better a robot's performance, the more confidence humans have towards it, and the more reliant they feel upon automation.

We note in particular that the revised H1 leaves out robot performance and confidence level, and H2 leaves out the difficulty of the task and the frequency of task delegation. Stated otherwise, the totality of our findings seems to suggest that high confidence in a robot's ability does not necessarily translate to robot-reliant behavior in a critical situation, nor does the lack of confidence in a robot necessarily result in self-reliant behavior.

What may have caused these apparent discrepancies would be an interesting area for future work, but we do wish to propose a tentative conjecture. As part of our instructions to each

[12] Robinette, et. al. "Overtrust of Robots" (See footnote 5).

subject, we went into extensive details regarding the catastrophic nature of the emergency, humanized the victim via giving her a faux name and biographical information, stressed that there was only one chance to attempt the mission (no retries were permitted), and advised strongly against taking unnecessary risks. We should also mention that there was a large gap in difficulty between the easy and the hard environments: the former had a sparse number of hazardous grids and did not require the respondents to make risky choices as long as they were minimally cautious, while the latter made it unavoidable for them to make numerous, highly risky choices that could spell instant failure. Given these stakes, we surmise that the high frequency of task delegation in a difficult environment has to do not so much with trusting the robot per se, but rather with a desire to avoid direct responsibility in the likely event of failure; on the other hand, the low frequency of delegation in an easy environment appears to be largely motivated not by a distrust of the robot, but a desire to take credit (if only in their own mind) for a life-saving task that the subjects felt capable of performing unassisted.

To verify the above conjecture and to supplement the shortcomings in this experiment, a future study might conduct similar sets of experiments, preferably in a real-life context and a larger group of subjects, with the nature of the tasks (i.e. critical vs. mundane) as one of the key independent variables. The dependent variables could include objective measurements of task delegation, as well as subjective measurements of confidence level, trust, and sentiment of reliance upon automation. Other independent variables such as task difficulty and/or autonomous performance could then be combined to look for possible interactions.

VI. Conclusions

We simulated an emergency search-and-rescue mission in a discrete 2-D environment comprised of two difficulty levels (easy vs. hard) and a robot whose autonomous performance was also manipulated (good vs. bad performance), in hopes of investigating their impact on humans' willingness to delegate critical control tasks to a robot. For dependent variables, we measured the participants' frequency of task delegation, along with the subjective measurements of their confidence level in the robot and their sentiments regarding the robot's reliability.

While we began the experiment with the assumption that all of the above-mentioned dependent variables would correlate strongly amongst themselves, this assumption turned out to have been misplaced. Our results indicated that in difficult environments, humans had strong feelings of reliance upon the robot and deferred control tasks regardless of the robot's prior performance, while in easy environments humans had weaker feelings of reliance and tended to perform the tasks on their own, again regardless of the robot's prior performance. While the performance metric did show a significant correlation with humans' confidence and trust in the robot, such trust did not correlate significantly with the frequency of actual task delegation.

Our conjecture as to the reason for the above is that, especially in critical situations, humans' behavior regarding task delegation to a robot may be driven more by their desire to avoid direct responsibility for failure and to gain credit for success, than by their trust in the robot's abilities. We believe this is a subject ripe for a future study.