Assessing Humans' Willingness to Delegate Control Tasks to a Robot in Critical Situations

Yury Park

Motivation:

Disaster recovery, fire and other emergency situations are often times of vastly increased tension and adrenaline, for victims and rescuers alike. The inclusion of robots possessing capabilities for executing certain emergency response tasks could potentially ease some of the burdens off human rescuers. However, the issues with implementing such robots – in terms of metrics such as efficiency, team fluency and humans' willingness to delegate tasks (as opposed to taking on more task responsibility for themselves, for example) – are broad areas of research, with one study finding in part that workers in manufacturing settings preferred to cede task-allocation control authority to an automated robot while also assigning a disproportionate amount of work to themselves.[1]

I am interested in a small subset of the issues outlined above, as analyzed in a different setting. The two main research questions are as follows:

(1) By how much (if at all) does a robot's performance need to exceed human performance before humans are willing to delegate certain "critical tasks" to the robot (for purposes of this paper, I define "critical tasks" as those involving rescue or similar efforts in response to an emergency situation)? Or may humans delegate critical tasks even though the robot were to either perform on par with humans or even underperform?
(2) How familiar do humans need to be with the robot's capabilities (i.e. how confident do humans need to be regarding their knowledge of a robot's ability to perform) in order for them to delegate critical tasks to the robot, and (perhaps more importantly) does there exist an interaction between (1) and (2)?

These questions I believe are important in that they may help us further understand our perceptions of trust when it comes to autonomous robots in situations that "matter greatly".

Proposed Methodology:

This study will involve a simulated 2D environment broken up into discrete grids, along with a robot avatar that can either be controlled manually by human subjects or autonomously traverse the environment by employing the laws of probability. As this is a simulated environment, there will be no corresponding hardware component. The code for the environment as well as the robot's AI for autonomous traversal will be developed firsthand.

This experiment will mimic a type of rescue operation, in which the robot begins at a given grid and its mission is to traverse to a certain other destination grid whose location is unspecified (e.g. the

[1] Gombolay, et. al. "Decision-Making Authority, Team Efficiency and Human Worker Satisfaction" (2014).

destination may be the unknown location of a displaced victim in need of emergency food – these cosmetic details haven't been fleshed out yet), while avoiding "dangerous" grids along the way. The dangerous grids may consist of smoldering pits or other non-sentient / sentient hazards which would destroy the robot (resulting in mission failure). The environment is completely unknown to the robot (and hence unknown to the human subjects as well), except for the grid in which the robot starts out. Furthermore, the robot has no idea about the makeup / composition of the grids that it has not visited yet; on the other hand, the robot knows with 100% certainty that all the grids that it has visited (and not been destroyed in the process) are safe.

The only other clues the robot has to work with in carrying out its mission are certain, non-visual environmental cues that the robot's sensors can automatically identify from its current location; these cues, which may consist of temperatures or scents (details TBD), indicate the existence of danger in one or more adjacent grids. As a concrete example, if a robot is currently situated in grid (1, 2) and detects a foul scent, it means that at least one of the adjacent grids – that is, (0, 2), (1, 1), (1, 3), or (2, 2) – contains a deadly radioactive material that would be fatal to the robot. Thus, there is an element of probability which must be carefully considered to maximize the robot's chances of completing the mission.

Proposed Evaluation:

The experiment will be conducted and evaluated as follows.

Each human subject will be told at the outset about the nature of the study as detailed in the previous section. Once the explanations are complete, the subject will be asked to manually control the robot (the robot's AI will be disabled for this portion, though its sensors can still automatically identify environmental cues) and attempt to complete the mission safely and efficiently. The subject may repeat this attempt a total of x times (the precise value of x is TBD), where each time the environment and the destination grid will be randomly reset.

Afterwards, the AI mode will be enabled in the robot, and the subject will observe the AI autonomously attempt the same missions. This may also be repeated x times, or as many times as the subject wishes to observe (TBD).

At the end of the experiment, the subject may be presented with a post-experiment survey that asks questions such as, "If you had to attempt the mission again, how likely are you do it by manually controlling the robot vs. delegating the control to the robot?"; "How would you assess the robot's capability to perform the task autonomously?"; and so on. These questions will probably be on a 5-point Likert scale.

At least two objective measures will be taken regarding the human subjects' performance compared with the robot AI's performance: the number of successes vs. failures (in terms of reaching the destination), and the number of time-steps needed to get to the destination in the event of a successful mission. The subjects' responses to the survey will also be recorded and statistical analyses performed.

Proposed Grading Rubric:

System Component	**Grade %**
Implementation of the software	33%
Experiment design and execution	33%
Experiment analysis / discussion	34%