Nudging or Waiting? Automatically Synthesized Robot Strategies for Evacuating Noncompliant Users in an Emergency Situation

Yuhan Hu
yh758@cornell.edu
Cornell University
Ithaca, NY, USA

Jin Ryu
jfr224@cornell.edu
Cornell University
Ithaca, NY, USA

David Gundana
dog4@cornell.edu
Cornell University
Ithaca, NY, USA

Kirstin H. Petersen
kirstin@cornell.edu
Cornell University
Ithaca, NY, USA

Hadas Kress-Gazit
hadaskg@cornell.edu
Cornell University
Ithaca, NY, USA

Guy Hoffman
hoffman@cornell.edu
Cornell University
Ithaca, NY, USA

ABSTRACT

Robots have the potential to assist in emergency evacuation tasks, but it is not clear how robots should behave to evacuate people who are not fully compliant, perhaps due to panic or other priorities in an emergency. In this paper, we compare two robot strategies: an actively nudging robot that initiates evacuation and pulls toward the exit and a passively waiting robot that stays around users and waits for instruction. Both strategies were automatically synthesized from a description of the desired behavior. We conduct a within-participant study ($n = 20$) in a simulated environment to compare the evacuation effectiveness between the two robot strategies. Our results indicate an advantage of the nudging robot for effective evacuation when being exposed to the evacuation scenario for the first time. The waiting robot results in lower efficiency, higher mental load, and more physical conflicts. However, participants like the waiting robots equally or slightly more when they repeat the evacuation scenario and are more familiar with the situation. Our qualitative analysis of the participants' feedback suggests several design implications for future emergency evacuation robots.

CCS CONCEPTS

• **Computer systems organization → Robotic autonomy**; • **Human-centered computing → Empirical studies in interaction design**; • **Hardware** → Analysis and design of emerging devices and systems.

KEYWORDS

evacuation robot, human-robot interaction, automatic synthesis, nudging robot, waiting robot, noncompliant user

ACM Reference Format:
Yuhan Hu, Jin Ryu, David Gundana, Kirstin H. Petersen, Hadas Kress-Gazit, and Guy Hoffman. 2023. Nudging or Waiting? Automatically Synthesized Robot Strategies for Evacuating Noncompliant Users in an Emergency Situation. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3568162.3576955

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03…$15.00
https://doi.org/10.1145/3568162.3576955

1 INTRODUCTION

What strategy should a robot take to lead humans out of a building during an emergency? Specifically, how should the robot act when evacuees are motivated to stay back and pick up valuable items, delaying their evacuation? We present a study that compares an actively nudging robot strategy, where the robot keeps pulling the human toward the exit, to a passive one, where the robot waits for the human to activate the evacuation when they are ready.

Robots could be helpful guides in emergency evacuation scenarios [20, 22, 23]. They can provide critical information, help people navigate around a building, and relay information back to mission control centers, all while helping people avoid a dangerous environment. That said, it is not clear what robot behaviours would be most effective and acceptable in assisting people to leave a building quickly and safely. In an ideal situation, the robot could simply lead the person out of the building. However, evacuees might not fully comply with a guiding robot. A major factor preventing rescue is pre-evacuation activities such as gathering valuables [3, 19, 24]. We are specifically interested in this scenario and simulate it by giving participants in our study conflicting goals of evacuating the building and collecting items.

In the study, participants move through a 3D-simulated smoke-filled environment, guided by an evacuation robot. The robot has a map-and-route display and knows the way to the exit. Participants can pick up items from rooms in the building for additional compensation but will lose all of the items if they do not evacuate on time. Each participant performs the evacuation task twice, with different exit locations and interacting with two different robot strategies: nudging and waiting.

The robots in our simulation are autonomous and operate using controllers that were automatically synthesized from high-level specifications (*e.g.*, "do not get closer than n units from the human"). Automatic control synthesis is still rare in HRI settings, and we highlight its utility in this work. The synthesis framework enabled us to clearly and easily specify the robot's two strategies. We were also able to readily change robot strategies when a previous attempt at this study failed to elicit the desired human-robot interaction.

From the quantitative analysis, We find the nudging strategy to be more effective and preferred by users in Round 1, when participants are unfamiliar with the evacuation process. However, we do not see a difference between strategies in Round 2, when participants repeat the evacuation scenario. A qualitative analysis of participants' comments sheds light on these findings and suggests design guidelines for emergency evacuation robots.

2 RELATED WORK

Researchers have studied evacuation incidents to understand human behavior in this high-stress situation [11]. Slow response and panic are major sources of delay and subsequent injuries [7, 26], with many of these delays caused by gathering valuables [3, 19, 24]. This suggests an opportunity to develop methods to manage evacuee movement, increasing the possibility of a safe escape.

Several HRI projects in recent years have studied robot-assisted evacuation, where robots serve as first responders and provide guidance. For example, Murphy et al. [15] highlighted the lessons learned from robots' assistance in search and rescue tasks, suggesting the importance of situation awareness and active information communication. Kim et al. [6] designed an evacuation guide robot for sensing environmental information and communicating between firefighters and people that need to evacuate. They suggest a portable, thermal-resistant robot design with multimedia communication. Martinson et al. [13] developed a cooperative human-robot fire-fighting team, where the robot communicated through a combination of speech and gesture, demonstrating a set of perceptual capabilities. Robinette et al. [21] conducted a physical evacuation study, finding participants overtrusted the robot.

Most similar to our work is a recent study by Nayyar et al. [17], which studied two robot evacuation methods, shepherding and hand-off, in a simulated environment. Their results revealed that the shepherding method was more effective in directing people toward the exit. Nayyar et al., as well as the other cited work, however, focused on experts or cooperative evacuees, under the assumption that these persons prioritize the evacuation of the emergency site. We, in contrast, are interested in robot interaction with humans who are not compliant during an evacuation, for example, when gathering valuables.

When it comes to how to best encourage the evacuation of noncompliant people, the robot can take an active or a passive role, and it is not clear which role is most appropriate for this scenario. On the one hand, humans are excellent at reasoning and planning in unstructured environments, even if they are not as precise as robots [16], suggesting a human-as-leader approach. On the other hand, humans are prone to emotional stress, hesitation and interruptions, which may affect their perceptions and decision-making [9], pointing toward a robot-as-leader approach.

In many examples of human-robot teaming, the human plays a leading role vis-à-vis a passive robot, for example, remote-controlling a robot swarm [8] or coordinating robots' motions [14]. Other work explored situations where a robot plays more of a leading role in a guidance situation [25]. Some researchers have offered a more nuanced view, discussing when a robot should take the lead in a human-robot collaborative team [2]. In this work, we compare an active "nudging" strategy with a more passive "waiting" behavior of a robot, comparing the resulting evacuation effectiveness and team dynamics.

3 METHOD

Our research question is: What robot strategy is most effective to evacuate noncompliant people in an emergency?

To study this question, we set up a simulated environment where participants have to evacuate a building within three minutes while being guided by a robot. Participants have two conflicting incentives: picking up valuable items and exiting the building. This is to simulate non-compliance based on the fact that people often delay evacuation due to them gathering valuables [3, 19, 24]. We run a within-participant design, where each participant experiences two robot strategies, nudging and waiting. The robots are autonomous; their control was automatically synthesized from high-level specifications (see: Section 3.4).

3.1 Study Design

Our study is a single factor (robot strategy), two-condition (nudging vs. waiting), within-participant design. We chose the two robot strategies as it is not clear who should have the leading role in the above-described evacuation scenario. On the one hand, humans may panic, prioritize valuables, and delay responses under stressful evacuation scenarios, whereas robots could be more rational in evaluating risks and planning a safe path. On the other hand, robots may not be able to sense essential considerations, such as the presence of victims and the user's emotional states.

Based on this dilemma, we developed two robot strategies: Under the **nudging** strategy, the robot takes the lead and constantly nudges the human toward the exit. Under the **waiting** strategy, the user decides when and where to go, and the robot provides help when requested.

More formally, in the **nudging** condition, the robot initiates evacuation from the beginning and always moves actively towards the exit. The robot only comes back toward the user when the user is out of sight. In the **waiting** condition, the robot waits for the user to initiate the evacuation sequence. Until then, the robot stays close to the user while informing them about the direction to the exit using its information display. In both strategies, the robot displays a localized map and arrows pointing to the exit (Figure 1c).

We ran a within-participant study, where each participant performs the evacuation task twice. Each round is guided by a robot with a different strategy and on a different route of similar complexity. The order of the strategies is counterbalanced.

Since we are interested in robot strategies for noncompliant behaviors, we added an item-collection task in addition to evacuation with a total of 40 items. Participants are rewarded based on the number of items collected upon evacuation success. To increase the tension between evacuating and gathering items as time runs out, the item bonus was added with an exponential growth scheme: $bonus = \$0.05 \times n\_items^2$. For example, for collecting the first item, a participant would receive \$0.05, but the tenth item is worth \$0.95. If participants have not evacuated within the time-bound, they will lose all of the item bonus.

Instead of testing a specific hypothesis, we use an exploratory methodology. Our intuition is that a nudging robot strategy would

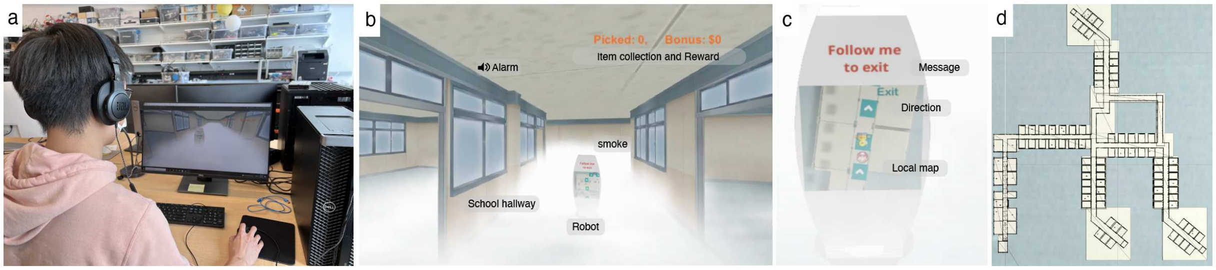

Figure 1: (a) Physical experiment setup (b) The simulator interface (c) The information display (d) The building's floor map.

be more effective and useful to evacuate non-compliant users, but we are unsure about the mechanism or effect size.

3.2 Participants

We recruited 20 participants via an internal university participant system (8 male, 11 female, 1 other, Age: $M = 21.65, SD = 2.70$). With each participant running two rounds of simulation, we gathered 40 data points in total. Participants received $15 in cash or course credit as compensation, plus a monetary bonus based on their experiment performance. The average number of items picked up was 6.88. The average resultant payout for the bonus was $3.06 per participant, with a maximum payout of $17.00 and a minimum of $0.00.

3.3 Experiment Platform: Unity

Since this study was initiated during COVID-19 gathering restrictions, and to be able to study a realistically complex environment, we created a virtual 3D evacuation task and simulated robots using the *Unity* game engine. The simulation renders a school hallway environment during a fire emergency, complete with smoke and a fire alarm (Figure 1b). The building is relatively complex (Figure 1d), so it is difficult for novel participants to navigate on their own. Each user performs the evacuation task taking a first-person view, controlling their character with a mouse and keyboard. A simulated robot (Figure 1c) has a cylindrical shape [5] of human height, displaying a message, a local map, and arrows pointing toward the exit.

In the first iteration of this study, we ran the experiment online with participants from *prolific.com*. However, we found significant latency in the game and variances among participants due to different hardware capabilities. As gathering restrictions eased, we eliminated such differences by conducting the experiment in a more controlled laboratory setup on a lab PC, as seen in Figure 1a.

3.4 Autonomy

The robots in our study are autonomous, using controllers that were automatically synthesized from high-level specifications [10]; that is, we described the desired behavior for the nudging and waiting robots using temporal logic and automatically generated the actuation commands for the robot in real-time, given its location, the time, and the human's location in the simulated environment.

In this work, we use Event-Based Signal Temporal Logic (STL) [4]. This formal language allows us to capture timing requirements, for example, "reach a location within 3 minutes", and goals and constraints over continuous robot behaviors and external events, for example, "stay within 2 meters of the human until the human presses the evacuation button", or "always stay at least 1 meter away from the human".

Using the framework of [4], we automatically convert the specification into an optimization problem which we solve at each time step of the simulation. This optimization problem takes as input the current location of the robot and the human, the specification, the status of external events, and the time, and returns a control input in the form of a velocity command for the robot to satisfy the task. If the optimization fails to find a feasible solution, meaning that the robot cannot guarantee the specification can be met, the robot stops its motion. The optimization may fail to find a solution because the specifications depend on the human's behavior—if the participant runs away from the robot, the robot may not be able to keep the desired distance from them. In such cases, we provide information as to why the robot failed to achieve its task and log it for further analysis. If the reason for failure is eliminated (in the example, the human returns to the robot's proximity), the controller recovers and continues to act according to the specifications. In the study, we did not observe any unrecoverable failure of the robot.

To define the specifications of the two behaviors, we went through several iterations, including an initial study involving 30 participants. In the first study, our specifications did not result in robot behaviors that were noticeably different for participants. We then tuned the parameters of distance and timing in the two robot controllers to serve their roles and gathered users' feedback for validation. Section 6.3 presents details on the previous iterations of this study. The specifications used in the current study are listed below.

Nudging Robot Specification: At the beginning of the experiment the robot must navigate to within 4 units of the person within 7 seconds. For reference, the environment is contained in a 187 by 216 unit rectangle, and the width of a hallway is 5.5 units. The robot must eventually be within 2 units from the exit of the environment within 10 minutes (the robot attempts to get to the exit as soon as possible, but we provide a larger time bound to ensure the optimization problem is always feasible). If the robot loses line of sight with the person, it must navigate back to within 10 units of the person within 160 seconds. Line of sight is lost when there is a wall between the person and the robot or if the robot is more than 40 units away from the person. During the entire experiment, the robot must avoid collisions and maintain at least one unit of distance from the person and at least 0.4 units from the nearest wall. For all tasks that eventually must happen, the robot will attempt to satisfy it as quickly as possible and prioritize tasks that have less time remaining to complete. For example, if the robot is navigating to the exit and loses line of sight with the human, it will immediately navigate back to the human to re-establish the line of sight and then proceed to travel to the exit as quickly as possible.

Waiting Robot Specification: Any time the robot is more than 4 units away from the person, the robot will, within 60 seconds, navigate to be within 4 units of the person. By doing this, the waiting robot will follow the person as they navigate through the environment. Once the person activates the evacuation signal, the robot will attempt to navigate to be within 2 units of the exit within 10 minutes of the signal. However, the robot must still remain within 4 units of the person. This results in the robot attempting to make progress toward the exit while ensuring the person remains close and is following. Similar to the nudging robot specification, the robot must avoid collision with the person and walls.

3.5 Procedure

The experiment began by going through consent forms and introduction slides that explained the task, reward system, game operation, and robot functions. To familiarize participants with the simulator, they started with a practice session. They tried out the interface on the experiment PC in a practice environment, navigating through hallways and picking up items. Participants were allowed to play with the simulation for as long as they wanted and ask any questions until they were ready to continue. Each participant played the evacuation game for two rounds, with two different robots on two different routes. The order of the robot strategies was randomized among participants.

Following the practice period, a researcher started the experiment simulation and verbally asked the participant to think-aloud during the game and asked again whether there were further questions. In each round, participants had three minutes to collect items and evacuate the building with the help of an evacuation robot. Participants wore headphones that played a fire alarm for the last minute of the game. Participants started in a school corridor and saw the robot approaching them. In the top right corner of the interface, participants could see the number of items collected and the amount of bonus money earned (Figure 1b). Once the participant reached the exit or reached the three-minute mark, the simulation ended, and we recorded a log file of the participant's trajectory, a video file of the screen capture, and the audio recording.

After each round, participants were asked to fill out a questionnaire through the *Qualtrics* online survey system to report their game experience and impression of the robot. A researcher then presented the instruction slides for Round 2 and repeated the above-mentioned process. After finishing Round 2, the participants filled out another questionnaire about Round 2, followed by several post-experiment questions to compare the two rounds.

3.6 Measures

We use both quantitative and qualitative metrics as dependent measures. As objective metrics, we measure **evacuation success**, *i.e.*, whether the participant reaches the exit within the time limit; **completion time**, *i.e.*, the time it takes to reach the exit if successful; and **distance from the exit at each time step**, which is calculated based on the participants' location, the exit location, and the shortest feasible route between the two. We also measure **the timing of participants' pressing of the evacuation button** (labeled *PressE Timing* in our graphs) when interacting with the waiting robot.

As subjective metrics, we designed a questionnaire to measure participants' perceived game experience (**Performance Satisfaction, Evacuation Prioritization, Goal Dedication, Stress Level, Difficulty, Immersiveness, Operation easiness**) and impression on the robots (**Helpfulness, Easiness, Clarity, Trust, Conflict, Intelligence, Willingness to follow, and Willingness to use in a real scenario**). Participants rate the items on a 5-point Likert scale ranging from 1 (*Not at all*) to 5 (*Completely*). To collect qualitative feedback, we include several open-ended questions to gather comments about the robot, the reasons for participants' compliant or noncompliant behaviors, and comparison of the two robots. We also encouraged participants to think aloud during the game, recorded the audio, and then transcribed the participants' speech and actions.

4 QUANTITATIVE RESULTS

We applied mixed methods to the data, including both quantitative and qualitative analysis. For the quantitative analysis, we ran t-tests on the evacuation performance and questionnaire metrics with robot strategy and round number as independent variables. We found increased performance and satisfaction in Round 2. Robot strategy had a different effect in Round 1 and Round 2: while a nudging robot is preferred in Round 1, a waiting robot is as good or slightly better in Round 2.

4.1 Order Effect in Evacuation Tasks

We found a strong order effect on many of our dependent variables. First and foremost, Round 1 had a much lower success rate (10%) compared to Round 2 (85%). 15 out of 20 participants improved their performance in Round 2 from failure to success. To evaluate the order effect, we ran paired t-tests on subjective ratings. Participants rated their **Satisfaction**, **Evacuation Prioritization**, robot usage **Easiness**, and **Clarity** much higher in Round 2 than in Round 1. We did not find a big difference in other subjective variables.

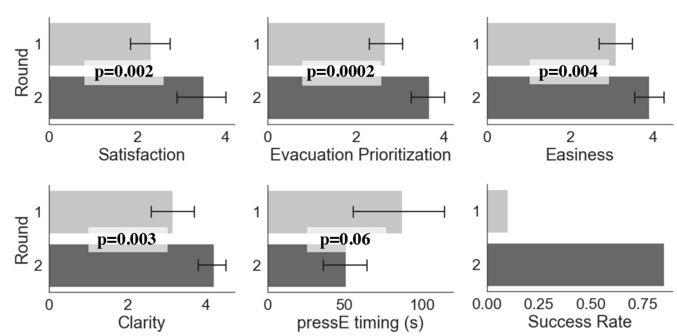

Figure 2: Comparison of subjective metrics, pressE timing and success rate between Round 1 and Round 2. The error bars show 95% confidence intervalss.

In addition, we ran an independent t-test for the timing of pressing the evacuation command for the waiting robot. **Users pressed the evacuation command noticeably sooner in Round**

2 ($M = 50.6s, SD = 24.8$) than in Round 1 ($M = 87.3s, SD = 51.5$), $t(13) = -2.03, p = 0.06$.

Overall, participants improved their evacuation performance in the second round, possibly because they prioritized evacuation more, found the robot easier to use with practice and started following the robot sooner.

4.2 Nudging Versus Waiting Robots for Effective Evacuation

To better understand the attitudes towards the robots while controlling for the strong order effects, we ran independent t-tests between nudging and waiting robots, separately for Round 1 and Round 2. **In Round 1, nudging robots were rated higher than waiting robots on Helpfulness, Trust, Clarity, Easiness, and less Conflict**. Interestingly, **in Round 2, participants perceived the waiting robot equally or slightly better than the nudging robot** on those scales (see: Figure 3). We did not find a strong difference in evacuation time or success rate between the two robots.

To better understand the differences between nudging and waiting robots, we calculated the distances from the exit along the routes, and visualized the participants' average distances from the exit over time (divided by the initial distance), separately for nudging and waiting robots (Figure 4). We also ran independent t-tests on the distances between nudging and waiting robots at three timestamps: $t = 60s, t = 120s, t = 180s$, presented in Table 1.

In Round 1, participants with a nudging robot started off moving faster toward the exit, getting ahead of the participants with a waiting robot. In Round 1, this trend continued until the end of the game. Although 90% of the participants failed the evacuation, participants guided by the nudging robot ended up closer to the exit than those guided by the waiting robot.

In Round 2, similar to Round 1, participants led by the nudging robot got ahead of the waiting robots at the beginning. However, the waiting group was as close to the exit as the nudging group at $t = 120s$. Most participants (85%) ended up evacuating within 180s, with no noticeable difference between nudging or waiting robots.

Round	Time	Average Distance / Initial Distance Mean (SD) %		t(16)	p
		Nudging	Waiting		
1	t=60s	80.23 (8.69)	89.15 (7.60)	2.44	**0.025**
	t=120s	52.81 (14.63)	66.83 (15.25)	2.10	**0.050**
	t=180s	17.46 (17.26)	35.98 (25.68)	2.09	**0.077**
2	t=60s	66.61 (13.51)	79.81 (11.77)	2.33	**0.032**
	t=120s	32.17 (24.14)	33.89 (17.23)	0.18	0.857
	t=180s	7.30 (16.15)	2.35 (4.41)	-0.93	0.37

Table 1: T-test results on participants' distances from the exit between nudging and waiting robots at three timestamps.

5 QUALITATIVE ANALYSIS

To better understand the reasons behind our quantitative findings, we used thematic analysis [1] on the participants' think-aloud recordings during the game and open-ended questions in the questionnaires. We first transcribed video and audio recordings of the evacuation game, focusing on participants' think-aloud quotations and interaction behaviors. Two authors independently reviewed all transcripts and highlighted the key observations for each participant. Two authors then collaboratively affinity-diagrammed these codings to identify repeated themes. After generating the mutually-agreed themes, two authors performed a thorough check of every quotation under each theme for accuracy and validation.

5.1 Theme 1: Nudging vs. Waiting

The first theme relates to participants' different perception of the nudging vs. waiting robot. Nudging robots were preferred for their efficiency and clear purpose, but were perceived to be slow. Waiting robots, on the other hand, caused higher conflict and mental effort.

5.1.1 Nudging robots. We identified positive comments from 12 (60%) participants and negative comments from 7 (35%) participants toward nudging robots.

Seven participants (35%) stated that they were more willing to follow the nudging robot because they **felt pressure to catch up with it**, *e.g.*, "The [nudging] robot forced me to follow it because I was afraid of losing it." (P8). When asked about the robot in the post-experiment questionnaires, five participants (25%) said **the nudging robot was efficient, "easy to follow" (P14), and "had a clear purpose" (P2)**. However, five participants (25%) perceived the nudging robot "**moved slowly**" (P13) and "[took] a longer path" (P4), possibly because it stayed further away from the user.

5.1.2 Waiting robots. 17 participants (85%) had negative comments, and 6 (30%) had positive comments toward the waiting robot. The leading comments revealed that **participants experienced more conflicts and added mental load when interacting with the waiting robot**, especially the following mode of the robot. 11 participants (55%) said the waiting robot blocked their way and trapped them as they were exiting the rooms. Six participants (30%) thought **the following mode was useless and inefficient**, *e.g.*, "I don't like the (following mode) because I don't need it to follow me. It can figure out what it's supposed to be doing." (P2)

Three participants stated that **having the option to switch modes took additional time, stress, and mental effort**. P17, for example, said: "You already have to think about the time you have left and to balance it with getting as many kits as you can. Adding another thing to think about doesn't help."

Still, three participants **enjoyed the freedom to choose prioritization** and lead the robot, *e.g.*, P19: "The [waiting] robot is better since I can choose my preference whether the robot is in front of me or behind me."

5.2 Theme 2: Compliance vs. Noncompliance

The second theme relates to participants' reasoning for their compliant and noncompliant behaviors while being guided by the robots. While most participants managed to cooperate with robots under conflicting interests, there were conflicts and confusion due to miscommunication, mistrust, and different physical capabilities.

5.2.1 ***Compliance****.* One major cause of compliance was the participants' **lack of confidence**. Fifteen participants (75%) indicated they had little knowledge of the building and were panicked and unconfident. *e.g.*, P16 said: "I trusted it certainly would be better

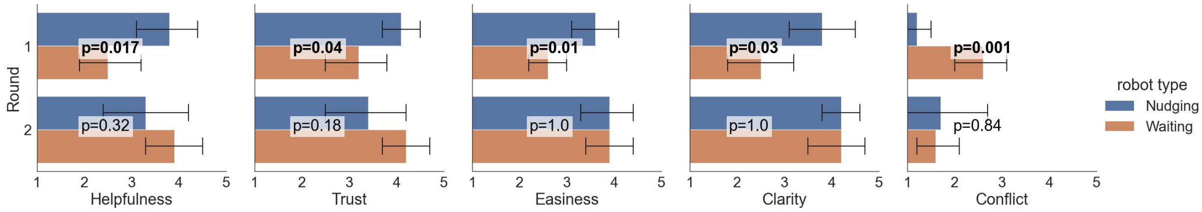

Figure 3: Subjective ratings on nudging and waiting robots of the two rounds. The error bars depict 95% confidence intervals.

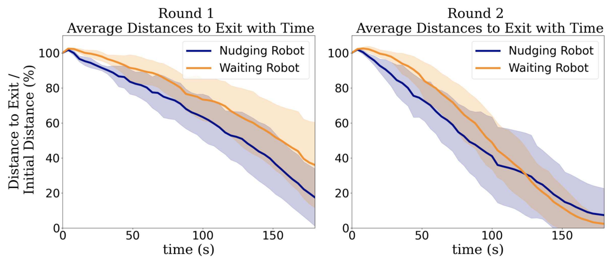

Figure 4: Participants' average distances from the exit as a function of time, guided by robot strategy. Shaded areas are standard deviations. In Round 1, the average distance is always lower with the nudging robot. In Round 2, participants with the waiting robot catch up around t=120s and end up closer to the exit.

than my own having no idea whatsoever of the layout of the building." Some participants said they followed the robot because there were no alternative behaviors they could take, including no item to pick up, no alternative route to explore, or lack of resources for independent evacuation (prior knowledge of the map or exit sign).

Eight participants (40%) said they followed the robot because they **trusted the robot**. Four participants (20%) mentioned using the robot's guide function while searching the room along their way, **trying to satisfy both goals**. Some participants inherently prioritized evacuation over item collection. P17, for example, said: "Because the ultimate task is still to make sure I get out of here. And less so like me trying to get a lot of items. Cause if I can't get out of here me getting the items is kind of useless."

Interestingly, fourteen participants' (70%) **compliance level significantly increased when users perceived higher risk**, such as siren breakout, distant robot, or smokier environment, *e.g.*, "I recognize there's definitely getting smokier[1], so I'm just gonna go really fast and stop looking for items." (P2)

5.2.2 ***Noncompliance.*** We identified five major reasons causing participants' noncompliance with the robot's guidance:

Seeing the Exit: Thirteen participants (65%) including those who have been compliant throughout the game turned noncompliant once they saw the exit for the first time. This implies that users may become noncompliant as their confidence and knowledge about the environment increases.

Physical Conflicts: With limited physical capabilities, robots could be slow and block participants' way, sometimes leading to a conflict point when the participants decided to walk ahead of the robot, suggested by notes from nine participants (45%), *e.g.*, "at one point, the robot stopped while I went on ahead. I ended up panicking and choosing my own direction." (P7)

Trust: Eight participants (40%) mentioned that they were skeptical of the robot's guide, especially when it took longer than expected to reach the exit or when the robot behaved "uncertain". Two participants left the robot for another corridor due to mistrust. For example, P2 said: "I don't know if it's leading me towards the fire. I also don't know why I judged the robot but I did."

Confusion: Six participants (30%) expressed confusion with the robot's motion, such as moving back and forth, standing still to wait for users, and circulating. The confusion about robot movements could also lead to trust breakdown as mentioned above.

Goal conflicts: Three participants (15%) ignored the robots because they prioritized picking up items over evacuating, especially at the early stage of the game.

Despite noncompliance, participants found their own ways to reconcile with the robots after the conflict. For example, when being blocked by the robot, some users got close to the robot, forcing it to move back to satisfy its specification, while some redirected the robot by making it follow the user in different directions. When people lost their robot, some stood still and waited for the robot to return, some reconciled upon the re-encounter with the robot after some self-exploration. The others actively went back to find the robot.

5.3 Theme 3: Robot Design Parameters

The third theme elaborates the relationship between robot design parameters and evacuation effectiveness; we identified three relevant parameters: **proximity, speed, and information display**. We found 26 quotes from 10 participants corresponding to robot's proximity, 42 quotes from 10 participants commenting on speed, and 35 quotes from 11 participants about the information display.

5.3.1 ***Proximity.*** The qualitative data revealed **how close the robot stayed to the user affected users' mental stress, willingness to follow, the robot's perceived speed and ultimately the compliance of the evacuation task.** Several participants reported feeling stressed and having the pressure to catch up with the nudging robot as it got farther ahead.

Having the robot stay closer to the users added to its perceived speed and likability, as the robot seemed to wait for the user and respond more quickly. For example, P17 said: "This (waiting) robot stops like way quicker than the other one; this would stop as soon as you have left it."

[1]In reality the smoke level did not change throughout the experiment or the building.

5.3.2 ***Speed****.* Although the robots' maximum speed is identical to the humans' in the simulation environment, the robot was perceived as "slow" by at least seven (35%) participants, probably due to the robot's collision avoidance specification. **Because of the slower perceived speed, participants have lost patience, affecting their willingness to comply, and increasing stress levels.**

Interestingly, two participants perceived the robot to be faster in the second round of the game, regardless of the type of robot, even though the maximum speed of the two robots was set the same.

5.3.3 ***Information Display****.* Seven participants (35%) mentioned that **the map display on the robot was helpful, easy to use, and increased its believability,** *e.g.*, "(the blue arrows) made me feel like a little bit more certain about the robot's path." (P2)

However, five participants (25%) said **reading the map was difficult** as the display was small when further away, and there was too much information to process in a stressful situation.

Seven participants (35%) **wished for more information** from the robot, such as the distance from the exit, the time left, and the locations of items. In addition, the robot could **communicate more explanation of its own movements**. Six participants (30%) mentioned their confusion regarding the robot's movements, such as when the robot paused or moved in circles to avoid obstacles, calculate its paths, or when there was not enough time left for evacuation, *e.g.*, "When I first switched the mode, I was unsure if I did it right because the robot did not immediately move in front of me." (P15)

5.4 Theme 4: Emotions and Stress

Theme 4 relates to participants' emotions under evacuation pressure. We found that participants were experiencing intense emotions and uncertainties under the evacuation scenario.

Eleven participants (55%) expressed their **stressful and frustrated** feelings. P18, for example, said after the alarm went off: "Does it mean I'm dying? There's no use. I'm dying anyways."

Similarly, eight participants (40%) were stressed about **the sense of uncertainty**, wondering about the size of the building, distance from the exit, liability of the robot, accessibility of the items, and even the feasibility of the task, *e.g.*, P6: "Is this the right way? I'm putting a lot of trust in you and I don't know if it was deserved."

Some participants couldn't make their own decisions under stress thus relied blindly on the robot's guidance. Four (20%) participants said they **"felt less stressed to have the robot present"** (P6).

At the same time, ten participants (50%) felt stressed about the robots' behaviors and were **less tolerant of robots' mistakes**, such as when they felt the robot was too slow or thought the robot was guiding toward the wrong direction. Some participants even felt betrayed by the robot; for example, P18 said: "Oh my god, it's a dead end! Oh I'm going to die (...) This is because of the robot. Where is the robot? I'll punch the robot!"

5.5 Theme 5: "Reasoning About Order Effect"

The fifth theme reasons about the order effects (Section 4.1), collecting different behaviors and mindsets in the two rounds. 32 quotes from 11 participants belonged to this theme, among which 8 described "regrets" of Round 1 and 24 noted the changes in Round 2.

In the first round, participants' actions were more ambitious with acquiring kits and informed by their own experience. Five participants (25%) said they **underestimated "how big the building was and how long it would take to get to the exit"** (P15).

Eight participants (40%) indicated that they "tried to be faster" and **prioritized evacuation more** in Round 2 realizing how hard it was to evacuate the building in Round 1. For example, P6 said: "The way I should've done the first time is I find the exit first and then go back." Four participants (20%) noted that they feel more comfortable and "**less risk**" (P2) in Round 2.

Five participants (25%) indicated their **trust changed** in robots because of their experience in Round 1. Some participants trusted the robot more: *e.g.*, "I realize last time I was a fool for not following the robot's lead." (P6); others were skeptical about the robot as they didn't lead them out in the first round.

6 DISCUSSION

In this section, we discuss the limitations of the study and present design implications and previous iterations of robot strategies.

6.1 Limitations

First, although we tried to make the simulation as realistic as possible, evacuation in a real-world environment with physical robots would be significantly different. For one, people's perceptions of proximity could change. In a physical setting, people might be more cautious of physical conflicts with the robot, thus having a larger safety zone than that in the simulator. Robots, on the other hand, have more physical constraints given their hardware capacities, such as maximum speed and turning radius. Generalizing our findings to physical robots would require additional testing and verification in physical environments.

We also observed different levels of operational skills among participants, which affected their task performance. Some participants exhibited awkward navigation controls, such as moving in a zigzag pattern or struggling to pick up items. Although we added a practice session with all the possible actions available, we could not entirely eliminate these differences. This possibly has to do with participants' familiarity with playing computer games. On the other hand, in a real-world emergency environment, we may also see differences in humans' moving capabilities, such as a slow motion for injured people or erratic behavior due to panic.

6.2 Design Implications

Our findings lead to the following design implications:

6.2.1 Nudging is Useful in an Unfamiliar Setting. We found a strong order effect and differences in the perception of the robots between the two rounds. In Round 1, most participants failed the evacuation, possibly because they underestimated the difficulty of the evacuation, spent more time gathering items, or were overwhelmed by environmental signals. In this case, the nudging robot strategy was especially useful for encouraging participants to evacuate as soon as possible and decreasing their mental loads. This indicates that giving users more control authority is not always beneficial, especially when users are in a new situation and stressed.

As a side note, despite observing a strong order effect, we did not add more training steps besides the practice session because we wanted to understand the impact of robot strategies on novice users. In real-world evacuation scenarios, users are most likely unfamiliar with the situation. Round 1 closely simulates the nature of emergency evacuation scenarios with unexpected environments and little prior practice.

6.2.2 Good Information Display Design is Critical. It is crucial to carefully design the information display of an emergency evacuation robot, such as by displaying the evacuation route and estimated time toward an exit as supporting proof. We found that users appreciated this information but that design issues with our information display made some users skeptical about the robot's decision. Having too much information requires time and effort for users to process, causing inefficiency and stress. Given the variances between participants' preferences, another idea is to customize the display based on participants' movements and emotions: for example, a noncompliant user may need more explanation to persuade them for evacuation, while a compliantly-following participant would only need the robot's own movements.

6.2.3 Stay Close By but Don't Interfere. Another variable that has affected evacuation effectiveness is the distance the robot stays from the user. While a close-staying robot may increase the sense of security and its likability, a distant robot may provide pressure for participants to catch up, thus encouraging evacuation. In addition, a robot nearby could result in more conflicts when hindering users' motions. Although a larger distance seems more favorable for evacuation purposes, a distant robot could cause trust issues as it becomes less visible when far away. In such cases, the robot may have to return to the users to persuade them to evacuate when observing trust breakdown signals.

6.2.4 Experts Need Less Nudging. Participants improved their evacuation performances significantly when given a second chance. This is consistent with findings from several evacuation studies where repeated exposure could improve participants' evacuation performances [12]. Interestingly, unlike Round 1, waiting robots were equally or even more liked than nudging robots in Round 2. This indicates that when participants have more knowledge of the task, robots, and environments, they are more comfortable taking control authority, making important decisions (*e.g.*, when to evacuate), and choosinng preferences of the robot's behaviors (*e.g.*, staying in front of or behind the user). This may shed light on the situation of evacuating experienced residents who have prior knowledge of the building and are familiar with the robots or emergency experts who are calm enough to make reasonable decisions. Having robots take a follower role in these cases could maximize outcomes by giving humans more flexibility in the evacuation task.

6.2.5 Pay Attention to External and Affective Signals. Besides the robot behavior, environmental signals and emotional state could also affect users' compliance level. In our study, participants became more compliant when they perceived higher risks and less compliant when they saw the exit. This suggests that a robot would have to monitor environmental signals such as smoke, noise level, and evacuation signage to account for users' behaviors and provide the best assistance. Stress and emotions have also affected participants' attitudes toward the robots and the evacuation performance. The design of an evacuation robot should carefully deal with users' emotions by monitoring their stress levels and developing different strategies to calm them and develop trust, such as by providing reassuring information and accompanying them with a physical presence.

6.3 Previous Iterations and Pilot Studies

The study presented here is a second attempt to answer our research question. We ran an almost identical experiment before the one reported here, and found that some aspects of the robot's strategies led to failures in human-robot interaction that were unrelated to our research question. Since we use control synthesis from high-level specification [4], we could quickly generate and test different robot behaviors between studies by simply changing the specification in a human-readable text file.

In our first study, we found that the difference between the two robot strategies, after the evacuation was initiated, was not noticeable by participants. Using Event-based STL, we could go through several iterations to tune the parameters in the robot specifications, including distances and timing, to address this issue.

Eventually, we specified two strategies: the nudging robot would travel as far away from the user as possible, as long as it is not out of sight of the human. The waiting robot would stay within a close distance and get back to the user as quickly as possible when it detects the user moving. To do so, the robot was tasked with *always* staying within a following range of the person. If the robot could not do this (the person moved faster than the robot could follow) the robot would stop moving because the specification was violated. We changed the task to be that the robot must *eventually* be within a following distance in a set time if the human ever moved further away. In this way, the task is only violated if the robot cannot re-establish the following distance in the allotted time.

To the best of our knowledge, this is the first study where automatically synthesized robot control is used in the service of an HRI study that is not about the specification language itself, as in, for example [18]. Using synthesized control over open-loop control or a Wizard-of-Oz methodology has two distinct advantages: It allows robot behavior that depends on the participant's actions while examining the robot's autonomous behavior with all its possible pitfalls, and it allows researchers to easily change, within a matter of minutes, the robot behaviors thus iterating quickly on possible behaviors. It would be interesting to consider the parameterized behaviors as a spectrum to vary using synthesized control for future robot design. Alternative evacuation behaviors may include a pushy robot that stays behind the user and pushes them toward the exit and a static robot that pauses at each crossroad and points toward the exit.

7 CONCLUSION

In this paper, we designed and compared two automatically synthesized robot strategies, nudging and waiting, for evacuating noncompliant users in a simulated emergency task. The results suggest that an actively nudging robot is more effective evacuating novice participants. Waiting robots could lead to physical conflicts and higher mental efforts, however, might be suitable for experts or experienced users. The quantitative and qualitative results provide several design suggestions for future emergency evacuation robots.

ACKNOWLEDGMENTS

This work was supported by the NSF Grant IIS-1830471.

REFERENCES

[1] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. SAGE Publications, Thousand Oaks, CA.

[2] Brenda Castro, Montana Roberts, Karla Mena, and Jim Boerkoel. 2017. Who takes the lead? Automated scheduling for human-robot teams. In *2017 AAAI Fall Symposium Series*. AAAI, 85–89.

[3] Robyn R. M. Gershon, Lori A. Magda, Halley E. M. Riley, and Martin F. Sherman. 2011. The World Trade Center evacuation study: Factors associated with initiation and length of time for evacuation. *Fire and Materials* 36, 5-6 (Feb. 2011), 481–500. https://doi.org/10.1002/fam.1080

[4] David Gundana and Hadas Kress-Gazit. 2021. Event-based signal temporal logic synthesis for single and multi-robot tasks. *IEEE Robotics and Automation Letters* 6, 2 (2021), 3687–3694.

[5] Jonathan Jaramillo, Andrew Lin, Emma Sung, Isabel Jane Hunt Richter, and Kirstin Petersen. 2021. Mobile, Inflatable Interface to Support Human Robot Interaction Studies. In *2021 18th International Conference on Ubiquitous Robots (UR)*. IEEE, Gangneung, Korea (South), 320–325.

[6] Young-Duk Kim, Yoon-Gu Kim, Seung-Hyun Lee, Jeong-Ho Kang, and Jinung An. 2009. Portable fire evacuation guide robot system. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, St. Louis, MO, USA, 2789–2794.

[7] Margrethe Kobes, Ira Helsloot, Bauke De Vries, and Jos G Post. 2010. Building safety and human behaviour in fire: A literature review. *Fire Safety Journal* 45, 1 (2010), 1–11.

[8] Andreas Kolling, Phillip Walker, Nilanjan Chakraborty, Katia Sycara, and Michael Lewis. 2015. Human interaction with robot swarms: A survey. *IEEE Transactions on Human-Machine Systems* 46, 1 (2015), 9–26.

[9] Kathleen M Kowalski-Trakofler, Charles Vaught, and Ted Scharf. 2003. Judgment and decision making under stress: an overview for emergency managers. *International Journal of Emergency Management* 3, 1 (2003), 278–289.

[10] Hadas Kress-Gazit, Morteza Lahijanian, and Vasumathi Raman. 2018. Synthesis for Robots: Guarantees and Feedback for Robot Behavior. *Annual Review of Control, Robotics, and Autonomous Systems* 1 (2018), 211–236.

[11] Erica D Kuligowski and Erica D Kuligowski. 2009. *The process of human behavior in fires*. US Department of Commerce, National Institute of Standards and Technology.

[12] Jing Lin, Lijun Cao, and Nan Li. 2019. Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology. *Advanced Engineering Informatics* 39 (2019), 53–61.

[13] Eric Martinson, Wallace E Lawson, Samuel Blisard, Anthony M Harrison, and J Greg Trafton. 2012. Fighting fires with human robot teams. In *IROS*. IEEE, Vilamoura-Algarve, Portugal, 2682–2683.

[14] Silvia Mastellone, Dušan M Stipanović, Christopher R Graunke, Koji A Intlekofer, and Mark W Spong. 2008. Formation control and collision avoidance for multi-agent non-holonomic systems: Theory and experiments. *The International Journal of Robotics Research* 27, 1 (2008), 107–126.

[15] Robin R Murphy and Jennifer L Burke. 2005. Up from the rubble: Lessons learned about HRI from search and rescue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 49. SAGE Publications, Sage CA: Los Angeles, CA, 437–441.

[16] Selma Musić and Sandra Hirche. 2017. Control sharing in human-robot team interaction. *Annual Reviews in Control* 44 (2017), 342–354.

[17] Mollik Nayyar and Alan R Wagner. 2019. Effective robot evacuation strategies in emergencies. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, New Delhi, India, 1–6.

[18] David Porfirio, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. 2018. Authoring and Verifying Human-Robot Interactions. In *The 31st Annual ACM Symposium on User Interface Software and Technology, UIST 2018, Berlin, Germany, October 14-17, 2018*. ACM, Berlin, Germany, 75–86.

[19] Guylène Proulx. 1995. Evacuation time and movement in apartment buildings. *Fire Safety Journal* 24, 3 (Jan. 1995), 229–246.

[20] Paul Robinette and Ayanna M Howard. 2011. Incorporating a model of human panic behavior for robotic-based emergency evacuation. In *2011 RO-MAN*. IEEE, Atlanta, GA, USA, 47–52.

[21] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, Christchurch, New Zealand, 101–108.

[22] Ibraheem Sakour and Huosheng Hu. 2017. Robot-assisted crowd evacuation under emergency situations: A survey. *Robotics* 6, 2 (2017), 8.

[23] Dylan A Shell and Maja J Matarić. 2005. Insights toward robot-assisted evacuation. *Advanced Robotics* 19, 8 (2005), 797–818.

[24] M.F. Sherman, M. Peyrot, L.A. Magda, and R.R.M. Gershon. 2011. Modeling pre-evacuation delay by evacuees in World Trade Center Towers 1 and 2 on September 11, 2001: A revisit using regression analysis. *Fire Safety Journal* 46, 7 (Oct. 2011), 414–424.

[25] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2010. A larger audience, please!—Encouraging people to listen to a guide robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 31–38.

[26] C Natalie van der Wal, Mark A Robinson, Wändi Bruine de Bruin, and Steven Gwynne. 2021. Evacuation behaviors and emergency communications: An analysis of real-world incident videos. *Safety science* 136 (2021), 105121.