IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Robot Navigation in Risky, Crowded Environments: Understanding Human Preferences

Aamodh Suresh[1], Angelique Taylor[2], Laurel D. Riek[3], Sonia Martínez[4]

Abstract—**The effective deployment of robots in risky and crowded environments (RCE) requires the specification of robot plans that are consistent with humans' behaviors. As is well known, humans perceive uncertainty and risk in a biased way, which can lead to a diversity of actions and expectations when interacting with others. To gain a better understanding of these behaviors, this work presents new data that aims to verify how these biases translate into a human navigational setting. More precisely, we conduct a novel study that recreates a COVID-19 pandemic grocery shopping scenario and asks participants to select among various paths with different levels of time-risk tradeoffs. The data shows that participants exhibit a variety of path preferences: from risky and urgent to safe and relaxed. To model users' decision making, we evaluate three popular risk models (Cumulative Prospect Theory (CPT), Conditional Value at Risk (CVAR), and Expected Risk (ER). We find that CPT captures people's decisions more accurately than CVaR and ER, corroborating previous theoretical results that CPT is more expressive and inclusive than CVaR and ER. We also find that people's self assessments of risk and time-urgency do not correlate with their path preferences in RCEs. Finally, we conduct thematic analysis of custom open-ended questions to gauge interest and preferences of navigational Explainable AI (XAI) in robots. We find a large majority of participants were interested in navigation XAI and want robots that infer how users plan paths in their environment. A large majority also showed interest in understanding robot's intention (path plans and decisions) through various modalities like speech, touchscreen and gestures. Several participants also expressed interest in learning the rationale behind robot's decision through high-level explanations. Our work provides crucial XAI design insights for deployment of robots in RCEs.**

Index Terms—**Human-Aware Motion Planning, Motion Control**

I. Introduction

ROBOTS are increasingly being deployed in everyday risky and crowded environments (RCE), including shopping malls, museums, streets, and sidewalks [1]. These environments contain multiple sources of risk (e.g., dynamic and chaotic human-motion trajectories) and uncertainty [2]. As robots become integrated in such environments, there is a need to design new safe and socially-acceptable navigation algorithms for them [3], [4], [5]. Precisely, a better understanding of how humans perceive risk and uncertainty in their decision making can help us design new robot motion strategies that are consistent with human's responses.

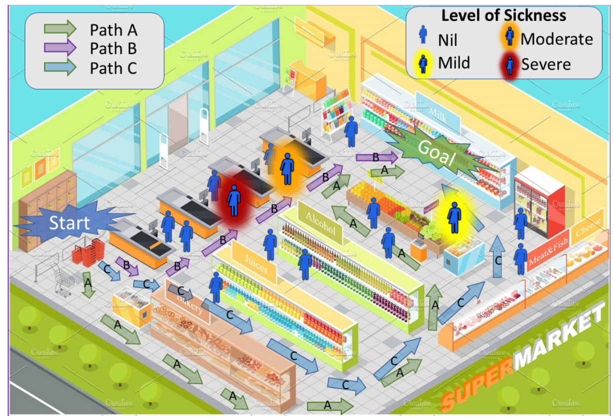


Fig. 1: Grocery store environment used in user studies. Participants select one of three paths (A, B, and C), to go from the entrance (shown as the 'Start') to the milk section (shown as the 'Goal'). The supermarket is crowded with people with levels of sickness ranging from 'Nil' to 'Severe'.

This is the case for autonomous vehicles transporting human passengers, or an autonomous shopping cart for the disabled. Here, the robot and human are co-located and the robot is required to understand the human's preferences in order to produce acceptable routes and actions. Similarly, a robot needs to understand how humans perceive risks in their vicinity in order to adapt its navigation around them.

In addition to robots understanding and generating human-like behavior, humans may also need to understand the robot's plans and rationale for clarity also known as Explainable Artificial Intelligence (XAI) [6], is the other key aspect of robot navigation in RCEs.

Motivated by this, our work validates commonly used risk models and standard questionnaires that evaluate human risk behavior against observed online decisions made by human participants in a risky and crowded grocery store environment. Additionally, we collect user data to understand human preferences for robot navigational XAI design through the lens of transparency, trust, information, and interpretability as highlighted in recent literature [6].

Prior Work: Humans portray various levels of rationality [7], [8] in decision making under risk and uncertainty. Thus, versatile models of risk perception [9] are useful to understand human navigation in RCEs. Available models range from those considering humans as completely rational (Expected risk (ER) behavior) and/or risk-averse (Conditional Value at Risk (CVaR) behavior [10]) to non-rational and possibly risk-insensitive (represented by Cumulative Prospect

Manuscript received: June, 23, 2023; Revised June, 24, 2023; Accepted June, 23, 2023.

This paper was recommended for publication by Editor Gentiane Venture upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by ONR N00014-19-1-2471, ARL W911NF-23-2-0009.

[1] A. Suresh is with the DEVCOM US Army Research Lab, Adelphi, MD, USA. E-mail: aamodh@gmail.com.

[2] A. Taylor is with the Department of Information Science, Cornell University, 2 W Loop Rd, New York, NY 10044, USA. E-mail: amt298@cornell.edu.

[3] L. Riek are with the Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA. E-mail: lriek@eng.ucsd.edu.

[4] S. Martínez is with the Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92093, USA. E-mail: soniamd@eng.ucsd.edu.

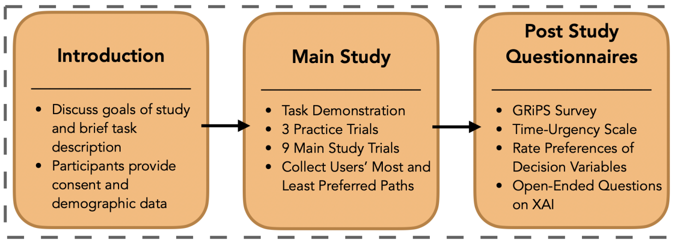


Fig. 2: An overview of our study design.

Theory (CPT) [11]). Previous work leveraged such models to pick routes [12] and generate reactive actions through direct adaptation [13], [14], or as starting points for navigation and learning algorithms [13], [15]. In theory, CPT has been shown to be the most expressive, versatile and inclusive [13], [14] model, thus capable of producing the widest range of behaviors among other models. However, little is known about the validity of these models in an RCE and how they compare with humans' self perception of risk.

Also, in practice, these approach is yet to be evaluated extensively in user studies pertaining navigation in RCE. To do so, user studies that employ natural or explainable metrics to humans need to be developed. Unfortunately, commonly used risk variables such as money [16], time [17], or collision probabilities [18], do not satisfy this criterion for all cases. In fact, recent studies have found that humans are often sub-optimal in planning paths in such situations [19]. These studies assume that the human is either "noisy-rational" or do not have correct environment models to choose optimally.

Other avenues of using risk for navigation RCE include fall risk assessment [20], [21], risk of localization and mapping [22], [23], and during search and rescue missions [24]. These arguments are from a robot's perspective which acts in an expected manner and also expects the human to do so. However, from a human-centered and XAI perspectives, the robot's "expected" behavior might lead to mistrust and confusion [25], [26]. To the best of our knowledge, general studies pertaining to everyday scenarios that employ more abstract cost interpretations are lacking, and are needed for better explainable AI design for robots in RCEs.

Contributions: In this work, through a novel large scale online study (n = 82), we bridge the gap in existing literature by characterizing, evaluating, and comparing human perception of risk in RCEs with risk models and standardized questionnaires. Furthermore, our contribution focuses on the comparison and validation of self-report measures of risks and human decision-making in risky scenarios that occur in real-world settings. To our knowledge, this is the first step towards robots that infer human perception of risk to use for intelligent robot decision-making in everyday settings. In addition, we provide new valuable insights for navigation XAI design, enabling more effective robot interaction with humans in RCEs. Specifically, our work addresses the following research questions:

RQ 1: What is the relationship between participants' path preferences and those arising from standard risk models?

RQ 2: What is the relationship between participants' self-risk, self-time-urgency perception and their path choices?

RQ 3: How do humans relatively weight time and risk to make navigational decisions in everyday scenarios?

RQ 4: What are the users' preferences to interact with robots navigating in everyday scenarios?

Our findings suggest that participants do not make decisions in an expected manner (in accordance with expected risk metric) and that CPT as a risk model captures the observed responses better than CVaR and ER. Interestingly, through the application of standard questionnaires, we find that there is a mismatch between humans' self-risk/self-time-urgency assessment and their actual choices. Additionally, participants generally give a higher weight to risk than time while choosing paths. We also obtain navigation XAI design for robots in RCEs. For example, we found that most participants want robots that can explain its rationale behind decision-making and they also suggested user interface design to have a two-way motion intention communication between users and robots. Thus, with the findings of this paper, the validated risk models can be used to plan paths, choose routes, generate reactive control actions, and learn navigation policies from observed human behavior as in previous work [13], [14], [15] in RCEs. Additionally, the XAI insights can be incorporated into robot design for efficient and meaningful interactions with humans in RCEs.

II. Methodology

We conducted an IRB-approved (approval code: 201638) within-subjects study employing the Qualtrics[1] survey platform. The main setting of the study, which aimed to evaluate people's risk perception in navigation, is inspired by the COVID-19 pandemic. This provides an easy-to-relate context for participants to think about decision making under risk. Thus, we consider a grocery-store shopping scenario, where the risk is characterized by being coughed at by potentially infected people. Participants were asked to imagine being an "Instacart[2] Shopper" who needs to go from the entrance of the store to the milk section. Time-urgency is characterized by the need to complete shopping quickly in order to get better ratings and tips.

This scenario is illustrated in Figure 1. Here, participants had three paths to choose from. Each path had a varying intensity of risk and time urgency (discussed in detail in Section II). The participants indicated their most and least preferred paths for each scenario. In the following paragraphs, we explain the scenario methodology and the list the post-study questions that we use (see Tables IIa and IIb).

Participants: We recruited 82 participants affiliated with a university campus through university list-serves and via word of mouth. The participants consisted of 27 females, 49 males, 1 binary/third gender and 5 that preferred not to answer this question. The ages ranged from 21-32 (mean = 25.6, SD = 2.5) and their educational background had a distribution of 68 in Engineering, 3 in Mathematics, 5 in Basic Sciences, 1 in Management, and 5 from other fields. Of non-STEM participants, there are 4 females, 2 males, fields include event planner/Management (1) and other (6).

Scenarios: Participants were presented with three paths to choose from, paths A, B, and C (see Figure 1). We used the situation of "being coughed at by sick people" to elicit risk for each path in every scenario. Risk was described by

[1]https://www.qualtrics.com/

[2]A grocery delivery service (`www.instacart.com`).

TABLE I: Description of decision variables and their ranges for each path in every scenario.

No.	Decision Variables	Range of values presented
$V1$	Time Taken	Path A : 20 mins Path B : 5 mins Path C : 10 mins
$V2$	Number of Sick people	Path A : 0-1 Path B : 2-3 Path C : 1-2
$V3$	Level of Sickness	0-3 for each path
$V4$	Chance of being coughed at	0-100 % for each path

three decision variables: "number of sick people" ($V2$), "level of sickness" ($V3$), and "chance of being coughed at" ($V4$) (see Table I). "Time taken" ($V1$) was used to elicit a sense of time urgency. While $V1$ varied from 5 to 20 minutes, $V2$ varied from 0 to 2, $V3$ from 0 to 3, $V4$ from $0-100\%$. To describe $V3$, we used the terminology: 0-Nil, 1-Mild, 2-Moderate and 3-Severe (see Figure 1). We purposefully kept the consequences of being exposed to a sick person abstract, in order to extract realistic risk perceptions from participants.

We administered nine trials with different values for decision variables, aimed at capturing a wide range of scenarios. In each trial, Path A was the longest and safest, path B was the shortest and riskiest, and path C had a length and risk that was in between. Participants chose their most and least preferred paths, thus providing a preference order. The risk variables for each trial are designed in such a way that the best expected choice[3] (A or B or C) varies across trials. We then group the nine trials into three levels of uncertainty (w.r.t. number of sick people). Comparing the chosen preferences with those based on expected risk, and across different levels of uncertainty, helps us understand how human choices are affected by uncertainty.

A. Measures and metrics employed

We obtained the following measures and metrics from each participant through nine trials and post-study questionnaires:

Path preferences: For each trial i, each participant revealed their path preference order by providing their most preferred path (MPP), denoted by m_i, and Least preferred path (LPP), l_i. Since by design, the shortest path (Path B) is the most risky and uncertain, while the longest path (Path A) is the least risky and uncertain, we define MPP and LPP as follows: $m_i, l_i \triangleq 0$, if participant chooses Path A for MPP or LPP; $m_i, l_i \triangleq 1$, for Path B, and $m_i, l_i \triangleq 0.5$, for path C. This definition conveniently encodes the level of risk and time-urgency$(0-1)$ numerically for each path.

1) Behaviors: The risk of each path (Fig 1) is described by decision variables $V2, V3, V4$ in Table I. While the variables $V2$ and $V3$ describe the path risk magnitude, $V4$ indicates the uncertainty associated with the path. Each path has an associated risk distribution. Different risk behaviors can lead to different path choices (MPP and LPP) for a given scenario. First, in the "expected behavior", the MPP m^{exp} and LPP l^{exp} are chosen in an expected manner, i.e. using Expected Risk (ER). Next, in the "risk aversion behavior", the MPP m^{av} and LPP l^{av} are chosen in a risk averse manner; i.e. considering worst case outcomes. Finally, in the "risk insensitive behavior", the MPP m^{ins} and LPP l^{ins} are chosen in a risk insensitive manner; i.e. not concerned about risk. We note that our previous theoretical results [13], [14] shows that CPT is the most inclusive model and captures all three perceptions. CVaR captures expected and risk averse perception, whereas ER only captures expected behavior. We provide the following example (a trial presented to participants in the study) for more clarity.

Example: Path A has 1 *moderately* sick person with a 55% chance of coughing at you. Path B has 1 *severely* and 2 *moderately* sick people, who each can cough at you with a 15% chance. Path C has 1 *severely* and 1 *moderately* sick people, who can cough at you with a 20% chance each. We then construct a risk distribution for each path following the valuations from table I[4]. The outcomes of Paths A,B,C are resp.: $[0,2]$, $[0,2,3,4,5,7]$ and $[0,2,3,5]$. The corresponding outcome probabilities are $[0.45, 0.55]$, $[0.614125, 0.21675, 0.108375, 0.019125, 0.03825, 0.003375]$, and $[0.64, 0.16, 0.16, 0.04]$, respectively. From a behavioral-risk perception perspective [13], [7], the valuation of these distributions would reflect the MPP and LPP choices. For "expected behavior", the expected risk for Paths A,B,C are: $1.1, 1.05, 1.0$ respectively. Thus, here $m^{\text{exp}} = 0.5$ (Path C) and $l^{\text{exp}} = 0$ (Path A). For "risk-averse" behavior, using either CVaR or CPT and setting the respective risk aversion parameters (q for CVaR, and high λ parameter for CPT; see [13]), results into $m^{\text{av}} = 0.0$ (Path A) and $l^{\text{av}} = 1.0$ (Path B). Lastly, the "risk-insensitive" behavior is obtained by considering a low CPT risk sensitive parameter (γ), thus producing risk valuations that are very low (≈ 0) for all paths. Since Path B is the shortest and Path A the longest, $m^{\text{ins}} = 1.0$ (Path B) and $l^{\text{ins}} = 0.0$ (Path A). •

The metrics that we use to analyze the data are the following. First, the *average MPP and LPP scores* are given by averaging over the trials as $\overline{M} = \text{avg}(m_1, \dots m_9)$ and $\overline{L} = \text{avg}(l_1, \dots l_9)$. Next, the *deviations from behaviors* are given by $J^{\text{exp}} = \text{avg}(|m_1 - m_1^{\text{exp}}|, \dots, |m_9 - m_9^{\text{exp}}|)$, for the expected-risk behavior, $J^{\text{ins}} = \text{avg}(|m_1 - m_1^{\text{ins}}|, \dots, |m_9 - m_9^{\text{ins}}|)$, for the risk-insensitive behavior, and $J^{\text{av}} = avg(|m_1 - m_1^{\text{av}}|, \dots, |m_9 - m_9^{\text{av}}|)$, for the risk-averse behavior. Finally, the deviations from RPMs are given as follows: For ER we consider $J^{\text{ER}} = J^{\text{exp}}$, for CVaR we take $J^{\text{CVaR}} = \min\{J^{\text{exp}}, J^{\text{av}}\}$, and for CPT we take $J^{\text{CPT}} = \min\{J^{\text{exp}}, J^{\text{av}}, J^{\text{ins}}\}$.

2) Self-reported measures: After collecting user path choices, we then conducted these post-study questionnaires.

GRiPS: General Risk Propensity Scale (GRiPS) [27] is a self-report measure (see Table IIa) which measures the participants' self risk-taking abilities through an 8-items questionnaire using Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). We denote the GRiPS questionnaire responses as $R \triangleq \{r_1, r_2, ..., r_8\}$ for the 8 questions.

Time Urgency: "Time Urgency Scale" [28] measures participants' self-reported assessment (Table IIb) of how urgent they think they behave, consisting of 6 questions (as commonly used [29]) using the same likert scale as above. We denote the responses as $T \triangleq \{t_1, t_2, ..., t_6\}$ for the 6 questions.

We employ metrics to evaluate the participants' self perception of risk and their observed risk-taking behavior, which areas follows: First, the *average Risk and Time Urgency scores* are given by averaging (and scaling between 0 and 1) over

[3]That is, the best according to the expected risk metric.

[4]The risk distribution is not shown to the participants in order to elicit natural scenarios and decision making

GRiPS Survey
1. Taking risks makes life more fun
2. My friends would say that I'm a risk taker
3. I enjoy taking risks in most aspects of my life
4. Taking risks is an important part of life
5. I commonly make risky decisions
6. I am a believer of taking chances
7. I would take a risk even if it meant I might get hurt
8. I am attracted, rather than scared, by risk

(a) The 8-item General Risk Propensity Scale (GRiPS) [27] Survey.

Time Urgency Survey
1. I find myself hurrying to get places even when there is plenty of time.
2. I often work slowly and leisurely.
3. People that know me well agree that I tend to do most things in a hurry.
4. I tend to be quick and energetic at work.
5. I often feel very pressed for time.
6. My spouse or a close friend would rate me as definitely relaxed and easy going.

(b) The 6-item Time Urgency Scale [28] Survey.

TABLE II: Post study Standardized Questionnaires

the trials as $\overline{R} = \text{avg}(r_1, \dots r_8)$ and $\overline{T} = \text{avg}(t_1, \dots t_6)$. So, a higher $\overline{R}, \overline{M}$ indicates respectively a more adventurous and time-urgent perception.

Next, we evaluate *Risk and Time Urgency similarity scores* to measure the deviation between users' self-risk or self-urgency perception and their risk or time-urgency level of their path choices. We define Risk similarity score $R^{sim} = \overline{M} - \overline{R}$ and Time-urgency similarity score, $T^{sim} = \overline{M} - \overline{T}$ and both $R^{sim}, T^{sim} \in [-1, 1]$. So a $R^{sim}, T^{sim} \approx 0$ indicates respectively the user's self-perception of risk or time-urgency is similar to the respective risk or urgency level of their path choices. A larger *positive* R^{sim}, T^{sim} value indicates users chose riskier and urgent paths than their corresponding self evaluation through the questionnaire. Similarly, a larger *negative* R^{sim}, T^{sim} value indicates users chose safer and longer paths respectively.

Decision Variable Preferences: After study trials and questionnaires were administered, we asked participants how they relatively weighed (as a %) each of the decision variables (Table I) in making their path choices to address RQ 3.

3) Open ended questions: We finally asked open-ended questions (listed below) to better understand human preferences towards designing robot navigation in RCE.

Q1: Would you like to know how robots make decisions and plan paths? If so, how do you want a robot to explain its thought processes behind its decisions and what modality (e.g., speech, expressions) would you prefer?

Q2: How do you want the robot to communicate its movement intentions (e.g., moving right or left)?

Q3: Would you like a robot to know how you are making decisions and planning paths? If so, how do you want to explain your intentions and what modality (e.g., speech, touchscreen) would you choose?

To summarize, the trial data helped us answer RQ 1. Question RQ 2 can be assessed from analyzing the trial data along with the questionnaire responses. The data on decision-variable preferences helped us answer RQ 3. Whereas the responses to open-ended questions helped us answer RQ 4.

B. Procedure

An overview of our study design is available in Figure 2. The study began with a consent form providing a brief description of the study. After giving informed consent, participants provided their demographic information including age, gender, occupation, and area of expertise. The main study then started, with a discussion the goal of the study, which is to investigate how people choose paths in risky situations. We further explained the user interface (UI) and elements of the study through a demonstration trial.

Participants then saw a demonstration scenario of the grocery store (Fig. 1), describing the four decision variables pictorially, in sentences and through a summary table[5] (see Table I). Then, they selected their most and least favorite paths. Based on their most preferred path choices, we randomly selected a risk outcome and display the final results[6] (e.g., 'You encountered no sick people').

Next, participants engaged in three practice rounds with three different scenarios and selected their most and least favorite paths. After the practice rounds, they then participated in the main study, which presented 9 different combinations of "risk" in each scenario. To remove ordering effects and the influence of regret, we randomized the ordering of the trials across all participants.

At the conclusion of the study, we administered the GRiPS [27] and Time-Urgency [28] questionnaires, to measure self risk-taking and time-urgency perception respectively among participants. We asked participants to weigh (as a percentage of) each of the four variables to make their path choices. Finally, we asked open-ended questions (listed in Section II-A) to better understand their preferences towards designing explainable AI for path planning in everyday scenarios.

C. Analysis

RQ1 Analysis: We examined the relationship between participants' path choices and the the corresponding choices that result from a risk model. To do this, we analyzed the descriptive statistics of the objective metrics that measure deviation from various behaviors and risk models.

RQ2 Analysis: We explored the relationship between participants self-perception of risk and time-urgency, compared to the expected (baseline) risk associated with their path choices in terms of MPPs and LPPs. We analyzed the descriptive statistics of the subjective metrics ($\overline{R}$ and $\overline{T}$) and the similarity scores (R^{sim} and T^{sim}). We performed correlation analysis to understand the interaction between self-perception of risk and time-urgency. We did a pairwise comparison of means (paired t-test) to reveal additional trends between participants choices and self perception.

RQ3 Analysis: We studied the relative importance that users give to each decision variable (in Table I) in order to choose their most preferred and least preferred paths. To do this, from this data, we created two new variables to measure the relative importance of time and risk used to make decisions. This was done by first averaging the three variables representing risk, and then expressing it as a percentage w.r.t. the total of time taken and average risk percentages. We provide descriptive statistics on the responses to post-study question on decision variable preferences.

RQ4 Analysis: As a second part to this study, we determined user preferences for XAI design through open-ended questions. Two members of our team performed thematic

[5]The decision variables were shortlisted and selected after many rounds of pilot studies involving different people.

[6]Through pilot studies we learned that displaying the trial results enhanced user engagement.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

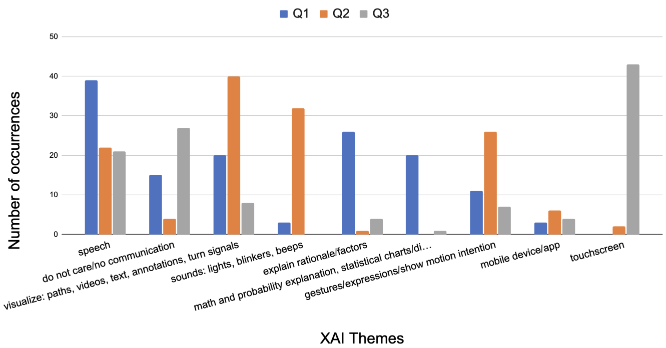


Fig. 3: This shows the number of occurrences of each explainable AI (XAI) theme identified in the data collected in responses to three open-ended questions from Section II-A to address RQ4.

TABLE III: Descriptive statistics of variables to compare users' decisions with standard risk-model decisions to address RQ 1.

Variable	Mean	95% confidence interval of mean	Standard Deviation	Range
$\overline{M}$	0.60	0.56 to 0.66	0.20	0.00 to 1.00
$\overline{L}$	0.40	0.35 to 0.47	0.30	0.00 to 0.94
J^{exp} for MPP	0.40	0.35 to 0.40	0.10	0.11 to 0.61
J^{av} for MPP	0.30	0.31 to 0.36	0.10	0.00 to 0.61
J^{ins} for MPP	0.30	0.28 to 0.36	0.20	0.00 to 0.89
J^{CVaR} for MPP	0.30	0.31 to 0.36	0.10	0.00 to 0.61
J^{CPT} for MPP	0.20	0.21 to 0.26	0.10	0.00 to 0.44

coding on open-ended question responses following grounded theory [30] as commonly done in the literature. This process involved reviewing responses (see Section II-A), generating high-level themes to capture key ideas in the data, reviewing the thematic codes with the team, negotiating them based on key ideas in the data, and repeating the process until all codes have been agreed upon. Next, we coded a total of 246 codes with our final set of codes shown in the x-axis of Figure 3. We computed the inter-rater agreement using Krippendorffs-Alpha as we used multiple codes for each quote. This is advantageous because it supports categorical and ordinal data. We found an IRR of 1.0 which is considered high agreement. We believe this is due to sparse responses with an average of 8.64 words per response, a minimum of 1, a maximum of 102, and a median of 4 words per response.

III. Results

We provide descriptive statistics of the relevant variables and metrics, including the mean, median, standard deviation, and 95% confidence interval. To study the correlation between two variables, we calculate the Pearson's correlation coefficient, along with null-hypothesis significance testing with threshold p-value = 0.05.

A. RQ1: Comparing users' risk perception

We provide descriptive statistics that compare risk models with users' decision-making (see Table III). Recall that MPP (similarly, LPP) for a j^{th} trial with $m_j = 0$ is path A with lowest risk and most leisured. Whereas $m_j = 1$ is path B with the highest risk and most time-urgency, and Path C is in between in both with a value of $m_j = 0.5$.

Path choice characteristics: From Table III, the mean, median, and confidence interval are over 0.5 for the average $\overline{M}$, and under 0.5 for the average $\overline{L}$. This indicates a preference towards Path B (more risky and time urgent), and a disinclination towards Path A (more safe and leisured). Additionally, $\overline{M}$ exhibits a full range from $0-1$, whereas $\overline{L}$ has a range $0-0.94$. Thus, although their preferences were scattered across the dataset, no participant disliked path B.

TABLE IV: Descriptive statistics of similarity scores of users' decisions in trials compared to their questionnaire responses to address RQ 2.

Variable	Mean	95% confidence interval of mean	Standard Deviation	Range
$\overline{R}$	0.6	0.55 to 0.64	0.2	0.00 to 1.00
$\overline{T}$	0.5	0.48 to 0.56	0.2	0.08 to 0.96
R^{sim} for MPP	0.10	-0.05 to 0.09	0.30	-0.84 to 0.94
R^{sim} for LPP	0.00	-0.08 to 0.07	0.30	-0.91 to 0.94
T^{sim} for MPP	0.10	0.04 to 0.15	0.30	-0.58 to 0.67
T^{sim} for LPP	0.10	0.00 to 0.14	0.30	-0.58 to 0.67

Behavioral Characteristics and model comparisons: The deviation from the expected behavior J^{exp} is in general greater than J^{av} and J^{ins}, indicating that expected behavior (from expected risk) is the least aligned with participants' preferences (see Table III). Also, from row 3, the min is > 0, indicating that not a single participant showed expected behavior across all their trials. The almost-similar statistics for risk-averse and risk-insensitive behaviors show that participants' exhibited both of these behaviors equally frequently. Also, the deviation J^{CPT} is the least, showing that CPT is a better model to approximate the participants' decision making. Similar deviation statistics were correspondingly obtained for LPP choices; hence, we omit its discussion here.

B. RQ2: Users' Self-Perception of Risk vs. Expected Risk

We provide descriptive statistics of the relevant variables (Table IV) and conduct correlation studies next. From Table IV, the statistics of the average survey responses $\overline{R}$ indicate that, in general, participants are more inclined towards taking risks, as the mean, median, and confidence interval are all over 3. However, a wide range and high standard deviation suggest a fairly diverse set of risk-taking behaviors. A similar trend is observed for time-urgency $\overline{T}$. To compare the participants' decision making characteristics with their self perception of time-urgency and risk, we first consider the similarity scores R^{sim} and T^{sim} (Table IV). The risk similarity score R^{sim} for MPP and LPP is balanced with the mean, median, and confidence interval close to 0, but have a high range and standard deviation. Hence, there are a variety of people with different perceptions, and the GRiPS responses may not fully represent their decision making in the study. A similar trend is observed for time urgency, with a slightly greater inclination of participants to act more urgently than what they indicate in the survey.

Next, we try to identify trends between the GRiPS and Time-Urgency survey responses and path choices by measuring correlation and performing linear regression. We highlight the significant results here. There was a significant interaction between the average survey responses for risk similarity score R^{sim} and time urgency similarity score T^{sim} for MPP and LPP with and $p < 0.05$ and a effect size <0.5. This implies that people who acted more/less riskier than they indicated in the GRiPS survey, also acted correspondingly more/less time urgent than they indicated in the time-urgency survey. This can arise because of the study construction, where paths which are shorter (time urgent) are also riskiest and vice versa.

Interestingly, there was an insignificant interaction between GRiPS responses $\overline{R}$ with MPP and LPP scores $\overline{M}/\overline{L}$ with p = 0.936 and p = 0.655, and effect size close to 0 revealing no relationship between them. This is consistent with prior research that shows people demonstrate cognitive bias during decision-making, relying on heuristics for decision-making [31]. Similar trends were found in terms of time-urgency. Thus, risk and time urgency measures may not be sufficient for social robot navigation scenarios.

Table V shows the paired t-test statistics of various pair-wise means comparisons. Interestingly we found significant evidence that $\overline{L}$ is greater than $\overline{R}$ (p $<$0.001) with a medium effect size (d $>$ 0.5). On average, $\overline{L}$ is greater than $\overline{R}$ by 0.18 (95% CI: 0.1, 0.26) units. By this, we can infer that participants had more disinclination towards riskier paths than their indicated risk appetite through the GRiPS survey.

C. RQ3: Users reliance on decision variables

We measured the relative importance that participants give to the four decision variables (Table I) to choose paths in terms of percentages. The data is described as boxplots (first 4 from the left) in Figure 4a. The last two boxplots represent the relative time and risk consideration, respectively. We note that the relative consideration of time has a mean 40.8% and median 42.8%, as opposed to relative consideration of risk which has mean 59.2% and median 57.2. So, in general, participants seem to consider risk factors more importantly than time factor while making decisions. However, with a large standard deviation of 30.2 and a full range of $0-100$ for both variables, the generalization may not apply to many participants. This again reflects the diversity regarding time and risk consideration for making path planning decision by humans. Thus, XAI needs flexible models in this regard.

D. RQ 4: Users' Interaction Preferences

We are interested in learning about users' preferences for XAI-capable robots in RCEs. To address our research questions qualitatively, we asked participants three open-ended questions from Section II-A to gauge their preferences w.r.t. how they would like robots to communicate their intentions. After conducting the analysis as described in Section II-C, we found many redundant responses for the three questions. We discuss descriptive statistics for each question and describe overarching themes we found in responses.

Descriptive statistics of open-ended questions: Figure II-A shows the descriptive statistics of the XAI themes used to code data across all open-ended questions. In summary, there are ten XAI themes identified in the data (x-axis of Figure II-C). From here, 73/82 participants indicated interest in XAI systems, and 8/82 participants did not want robots that explain their motion intentions. On the other hand, 69 out of 82 participants wanted robots that understand how users in their environment plan paths, 10 out of 82 do not, and 3 out of 82 participants indicated maybe. The 'touchscreen' theme achieved highest occurrence count. Several themes achieved the lowest occurrence score of 0 which include themes 'sounds, lights, blinkers, beeps' for how the robot knows what users want, 'math, probability, explanation, statistical charts/diagrams' for how to communicate intent, or 'touchscreen' in terms of how robots plan paths.

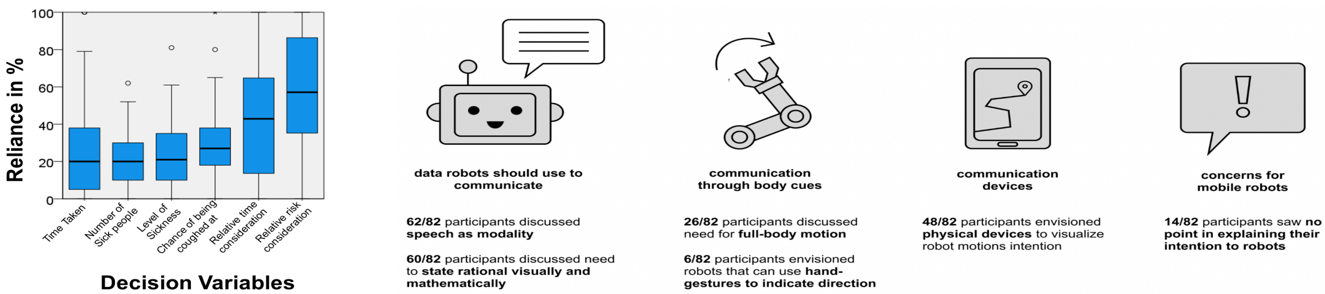
E. Thematic coding of Open-Ended Questions

We identified four overarching themes that inform users' preferences for XAI robotic navigational systems. These themes emphasize the importance of data variables used to interact with XAI systems for mobile platforms, how robots should communicate their navigation plans, and users' concerns about using XAI for mobile robots in everyday environments through the lens of transparency, trust, information, and interpretability [6]. Lack of effective mental models of robot decision making often leads to robot failures and lack of trust and safety around them. Also, explanatory, visual, or verbal information provide mechanisms for robots to express what data they are using for decision making to users. Furthermore, many "black-box" machine learning algorithms prevent users from interpreting how robots transfer inputs to outputs; thus, users need to interpret the inner workings of these algorithms to provide transparent feedback on how to improve robot decision-making processes.

1) What data robots should use to communicate with users: Users' identified a range of preferences for data robots can use to communicate its motion intentions. The most popular modality discussed by 62 out of 82 participants was speech, as it is convenient to communicate naturally. More specifically, 60 out of 82 participants discussed the need to state rationale visually such as with pie charts, using arrows, or visual representation of weights used in AI decision-making. Additionally, 26 out of 82 participants envisioned robots that can explain the factors behind their decision, consistent with prior work in XAI [32] e.g., level of sickness during COVID-19. Lastly, participants discussed the need for mobile robots that adapt to users' movement over time. P39 stated, "I would want to know what the robot is 'thinking' so that I know how to adjust my own behavior/path/location." We also found consistency with responses from STEM and non-STEM participants in this theme.

2) Robots that communicate motion intentions with body cues: Our analysis of the open-ended questions indicate that participants envision robots that communicate non-verbally using bodily cues. For instance, 26 out of 82 users discussed the need for full-body motion. They discussed robots that preemptively gesture in the direction they plan to move in before doing so. Also, 6 out of 82 participants envisioned robots that can provide hand gestures to indicate the direction they plan to move in. P84 said, "Something like a turn indicator on a car. It should flash/get attention visually and make a noise if possible so that people with low vision/hearing would be able to notice it". We found consistency with responses from STEM and non-STEM participants in this theme as well.

3) Preferred communication devices with robots: Our results show that users preferred several devices for communicating with robots. Participants envisioned intelligent user interfaces to facilitate interaction with robots inspired by Google Maps as a top-down map of the robot and people around it in real-time. Building on this, they discussed a feature that enables them to view a ranked list of paths the robot is considering and why it chose its current path in terms of factors. P9 said, "I would like to know the top options the robot considered and why it decided to go with its final choice." Another idea was for robots to have 'car-like' features like turn signaling with lights or using a human-actuated hand to point or gesture in the direction the robot intends to move in.



(a) Reliance on decision variables

(b) XAI Themes

Fig. 4: (a) Boxplots showing the users' preferences of decision variables (Table I) in percentage to choose paths in the trials to address RQ 3. (b) Themes that inform users' preferences for XAI systems in social navigation settings from Section II-A to address RQ 4

TABLE V: Paired t-test statistics of relevant variables to address RQ2

Variable X	Variable Y	P-Value	Effect Size (Cohen's d)	Difference Between Means (Y - X)	95% confidence interval of difference
$\overline{T}$	$\overline{R}$	0.029	0.249	0.070	0.010 to 0.140
T^{sim} for MPP	R^{sim} for MPP	0.029	0.249	-0.070	-0.140 to -0.010
$\overline{M}$	$\overline{R}$	0.544	0.068	-0.020	-0.090 to 0.050
$\overline{L}$	$\overline{R}$	<0.001	0.503	0.18	0.100 to 0.260
$\overline{M}$	$\overline{T}$	0.006	0.318	-0.090	-0.160 to -0.030
$\overline{L}$	$\overline{T}$	0.008	0.302	0.110	0.030 to 0.190

Lastly, 48 out of 82 participants envisioned physical devices to visualize robot motion intention information including on their phone or touchscreen devices. Non-STEM participants showed a variety of responses for this theme. Their feedback ranged from those that don't feel they should directly communicate with the robot to those that prefer robot communicate it's navigation intentions using touchscreen and speech.

4) Concerns about XAI for mobile robots: 14 out of 82 participants saw no point in explaining their intentions to robots. Instead, they envisioned robots that adapt to them. Furthermore, they wanted robots to be passive actors in their environment instead of an active agent that they can interact with, indicating "[...] this should be inferred. I don't want to change my natural behavior" -P21. One salient reflection resulting from the analysis is that some users expressed concern for robots using speech to communicate and they foresaw it as 'creeper' or 'annoying'. P9 said, "Speech is a convenient way [...], but it could get annoying if the robot is constantly talking about why it's going where it's going." Overall, there were a range of preferences identified in the data which highlights the need for systems with diverse XAI capabilities that adapt to users' preferences over time. Thus, robots equipped with more general models like CPT will be able to more effectively learn and infer human behavior [13].

IV. Discussion and Conclusions

We provide a brief discussion of our results, including implications and limitations of our work and possible avenues for future work and then finally conclude.

On average, the survey responses effectively captured participants' risk and time urgency behavior from study trials. However, these may not be fully indicative of people's risk and time-urgency preferences as there was a large standard deviation in the similarity scores. Participants tend to prefer safer paths the least compared to their indicated risk appetite for the survey. Furthermore, there was no significant correlation between participants' risk propensity and time-urgency.

Our study reveals that humans act in a diverse manner in RCE, thus motivating the need to equip robots with more inclusive risk perception models to foster co-agent interaction and integration in such environments. We also revealed that standard questionnaires that measure risk propensity and time-urgency scale of individuals may not be very reliable to consistently predict human decision making in everyday RCE. We believe that risk and time-urgency are subjective and circumstantial factors of human decision making. Thus, we feel a single generic risk and time-urgency questionnaire responses may not reflect the participants' decision making in all scenarios, as seen in our particular environment. Additionally, cognitive bias may also affect human decision-making [31], where participants rely on heuristics, thus making non-representative decisions w.r.t their beliefs. An in-person study and questionnaires focused on navigation in RCE might throw some more light on these interactions. Additionally, the robot may need to rely on other techniques such as online learning to understand the human perception of risk for efficient communication.

Our research demonstrates insights about users' preferences for XAI-capable mobile robots that communicate their motion intentions. Our findings suggest that robots using XAI modes of interaction can potentially improve HRI by sharing information such as sensor inputs and outputs, decision-making processes using ubiquitous devices (e.g., tablets), light indicators, and speech. Using these modalities, the XAI capbable robots can provide high-level information to interested pedestrians regarding its risk model choice with model weights and planned paths. This is in accordance with recent work in natural language processing [33] indicating humans find it useful for robots to provide high-level explanations of technical content.

Our findings also show that users' background strongly impacts their preference on modes of XAI from robots in terms of navigation tasks. In this way, participants with STEM backgrounds preferred mathematical explanations. However, non-STEM participants preferred robots to communicate using speech, without over-explaining their decision-making, and desired interpretability information from robots as they navigate in crowded environments (e.g., alert pedestrians it is coming in close proximity to them).

Limitations and Future Work: The benefit of conducting online studies was to elicit feedback from a large, diverse population. However, participants' path choices may not be fully representative of their actions in real-world settings or immersive experiences (e.g., virtual reality) where people walk at different proximities to the robots or play different roles during interactions. For example, users in a real-world shopping center can range from bystanders to those at further distances from the robot which may impact their preferences for robot navigation.We plan to expand our study to real-world settings in future work.

There might have been an inadvertent sampling bias resulting in a relatively younger age group of a student participant population, particularly from STEM backgrounds. Moreover, we conducted correlation analysis to understand the impact of age on MPP, LPP, and GriPS scores. We found no significant correlation between age and MPP and LPP score. Interestingly, we found a significant mild negative correlation between participants age and GRiPS score [r(82) = -.226, p = .041]. Thus, implying that people with higher age tended to think that they took lower risks in life as reported from the GRIPS survey. Future work with in real environments might help to better understand the path and risk preferences of users from the everyday public, different age groups, and diverse backgrounds.

Although, we had a large sample size in terms of number of people, we only collected limited data (only 9 trials) per participant to minimize fatigue. More data points are needed to explicitly compare and characterize decision making between participants and risk models. This limitation can be again alleviated by conducting in-person user studies where data can be collected in a natural and continuous manner (entire paths), which can then be used to perform more rigorous comparisons between various risk models.

Conclusions: In this work, we found that people tend to exhibit a variety of risk perceptions and behaviors in a crowded social navigation setting. We found that risk models like CPT, that are more expressive and inclusive, can better depict the observed human behavior, which thus support the previous theoretical findings. We also found that existing standard questionnaires to determine a users' risk-taking and time-urgency traits were not consistent with the exhibited behavior. In addition, we provided novel insights to consider for future XAI development for social navigation scenarios.

REFERENCES

[1] A. Taylor, S. Matsumoto, W. Xiao, and L. D. Riek, "Social navigation for mobile robots in the emergency department," *In Proc. of the IEEE Intern. Conf. on Robotics and Automation (ICRA)*, 2021.

[2] A. Taylor and L. D. Riek, "Regroup: A robot-centric group detection and tracking system," in *Proceedings of the 2022 ACM/IEEE Intern. Conf. on Human-Robot Interaction*, 2022.

[3] A. Taylor, S. Matsumoto, and L. D. Riek, "Situating robots in the emergency department," in *AAAI Spring Symposium on Applied AI in Healthcare: Safety, Community, and the Environment*, 2020.

[4] A. Taylor, M. Murakami, S. Kim, R. Chu, and L. Riek, "Hospitals of the future: Designing interactive robotic systems for resilient emergency departments," in *In Proc. of the ACM Conf. on Computer Supported Collaborative Work (CSCW)*, 2022.

[5] A. Suresh and S. Martínez, "Human-swarm interactions for formation control using interpreters," *International Journal of Control, Automation and Systems*, vol. 18, pp. 2131—-2144, 2020.

[6] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, "Explainable ai: A brief survey on history, research areas, approaches and challenges," in *CCF international conference on natural language processing and Chinese computing*. Springer, 2019, pp. 563–574.

[7] S. Dhami, *The Foundations of Behavioral Economic Analysis*. Oxford University press, 2016.

[8] S. S. Stevens, "On the psychophysical law," *Psychological Review*, vol. 64, no. 3, pp. 153–181, 1957.

[9] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 43–52.

[10] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, no. 1, pp. 21–42, 2000.

[11] A. Tversky and D. Kahneman, "Advances in Prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[12] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2020, pp. 43–52.

[13] A. Suresh and S. Martínez, "Planning under non-rational perception of uncertain spatial costs," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 4133–4140, 2021.

[14] ——, "Risk-Perception-Aware Control Design under Dynamic Spatial Risks," *IEEE Control Systems Letters*, vol. 6, pp. 1802 – 1807, 2022.

[15] K. Nar, L. J. Ratliff, and S. Sastry, "Learning prospect theory value function and reference point of a sequential decision maker," in *2017 IEEE 56th Annual Conf. on Decision and Control (CDC)*. IEEE, 2017, pp. 5770–5775.

[16] A. Majumdar and M. Pavone, "How should a robot assess risk? towards an axiomatic theory of risk in robotics," arXiv:1710.11040.

[17] S. Gao, E. Frejinger, and M. Ben-Akiva, "Adaptive route choices in risky traffic networks: A prospect theory approach," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 727–740, 2010.

[18] A. Hakobyan and I. Yang, "Wasserstein distributionally robust motion planning and control with safety constraints using conditional value-at-risk," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 490–496.

[19] S. Reddy, A. Dragan, and S. Levine, "Where do you think you're going?: Inferring beliefs about dynamics from behavior," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[20] R. S. Novin, A. Yazdani, A. Merryweather, and T. Hermans, "Risk-aware decision making in service robots to minimize risk of patient falls in hospitals," *arXiv preprint arXiv:2010.08124*, 2020.

[21] K. Koide and J. Miura, "Collision risk assessment via awareness estimation toward robotic attendant," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 011–11 016.

[22] O. A. Hafez, G. D. Arana, and M. Spenko, "Integrity risk-based model predictive control for mobile robots," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5793–5799.

[23] G. D. Arana, O. A. Hafez, M. Joerger, and M. Spenko, "Localization safety validation for autonomous robots," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 6276–6281.

[24] V. Shree, B. Asfora, R. Zheng, S. Hong, J. Banfi, and M. Campbell, "Exploiting natural language for efficient risk-aware multi-robot sar planning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3152–3159, 2021.

[25] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[26] B. Shneiderman, *Human-Centered AI*. Oxford University Press, 2022.

[27] D. C. Zhang, S. Highhouse, and C. D. Nye, "Development and validation of the general risk propensity scale (grips)," *Journal of Behavioral Decision Making*, vol. 32, no. 2, pp. 152–167, 2019.

[28] F. J. Landy, H. Rastegary, J. Thayer, and C. Colvin, "Time urgency: The construct and its measurement," *Journal of Applied Psychology*, vol. 76, no. 5, pp. 644–657, 1991.

[29] S. Mohammed and S. Nadkarni, "Temporal diversity and team performance: The moderating role of team temporal leadership," *Academy of Management Journal*, vol. 54, no. 3, pp. 489–508, 2011.

[30] V. Braun and V. Clarke, *Thematic analysis.* American Psychological Association, 2012.

[31] A. Bertrand, R. Belloum, J. R. Eagan, and W. Maxwell, "How cognitive biases affect xai-assisted decision-making: A systematic review," in *Proceedings of the 2022 AAAI/ACM Conf. on AI, ethics, and society*, 2022, pp. 78–91.

[32] G. Papagni and S. Koeszegi, "Understandable and trustworthy explainable robots: a sensemaking perspective," *Paladyn, Journal of Behavioral Robotics*, 2021.

[33] N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. F. Christiano, "Learning to summarize with human feedback," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3008–3021, 2020.