Beyond Omakase: Designing Shared Control for Navigation Robots with Blind People

Rie Kamikubo
College of Information
University of Maryland
College Park, Maryland, USA
IBM Research - Tokyo
Tokyo, Japan
rkamikub@umd.edu

Seita Kayukawa
IBM Research - Tokyo
Tokyo, Japan
seita.kayukawa@ibm.com

Yuka Kaniwa
Waseda University
Tokyo, Japan
ycrabring@ruri.waseda.jp

Allan Wang
Miraikan - The National Museum of
Emerging Science and Innovation
Tokyo, Japan
allan.wang@jst.go.jp

Hernisa Kacorri
College of Information
University of Maryland
College Park, Maryland, USA
hernisa@umd.edu

Hironobu Takagi
IBM Research - Tokyo
Tokyo, Japan
takagih@jp.ibm.com

Chieko Asakawa
IBM Research
Yorktown Heights, New York, USA
Miraikan - The National Museum of
Emerging Science and Innovation
Tokyo, Japan
chiekoa@us.ibm.com

Abstract

Autonomous navigation robots can increase the independence of blind people but often limit user control—following what is called in Japanese an "omakase" approach where decisions are left to the robot. This research investigates ways to enhance user control in social robot navigation, based on two studies conducted with blind participants. The first study, involving structured interviews (N=14), identified crowded spaces as key areas with significant social challenges. The second study (N=13) explored navigation tasks with an autonomous robot in these environments and identified design strategies across different modes of autonomy. Participants preferred an active role, termed the "boss" mode, where they managed crowd interactions, while the "monitor" mode helped them assess the environment, negotiate movements, and interact with the robot. These findings highlight the importance of shared control and user involvement for blind users, offering valuable insights for designing future social navigation robots.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; **Accessibility**.

Keywords

Autonomous robots, Social navigation, Crowded public spaces

With advancements in autonomous robot research and development, a growing number of robotic applications are being recognized, including those aimed for accessibility and social inclusion [35, 36, 67]. One actively explored use case is the development of navigation robots for blind people, with the hope to increase their independence and autonomy (*e.g.*, [4, 16, 33, 75]). However, as technical advancements push the boundaries for fully autonomous robots, blind users may find themselves as followers [89]. This dynamic can be described as *omakase* in Japanese [56], an analogy also used in a decision-making model for leaving decisions to others (*e.g.*, often to 'experts' in the situation [1, 72])—in this case, to the robot. In the context of disability, independence can be framed as control, choice, and power [65]. Such *omakase* robot interactions can create a critical design paradox for blind people, compromising independence when they have limited control over decisions about their environments [26, 44].

In this work, we aim to identify design strategies to enhance user control in robot navigation for blind people, with a focus on their choices of autonomy in complex real-world environments. While

recent research has explored user control beyond the omakase (or follower) model—such as active control as a *boss* [89] and companionship through *monitoring* surroundings [38]—there is limited understanding of how these choices play out in social navigation scenarios. These environments, which involve nuanced social factors with people around, can constrain robot performance [57] and make user control crucial for usability and overall acceptance [13]. This work explores desirable human-robot interactions that support greater user control and agency for blind people in socially complex navigation scenarios.

To start our exploration, we conducted a preliminary study involving structured interviews with 14 blind participants to understand their challenges in social navigation and the strategies they currently use. Building on these insights, we designed a two-part user study with 13 blind participants (including 3 overlaps), first having them complete navigation tasks with an autonomous robot in real-world crowded scenarios informed by our initial interviews. These tasks were intended to prompt participants to engage with different modes of autonomy, which we introduced by adopting design metaphors that were culturally relevant: following the robot's movements passively as Omakase Mode, requesting spatial information as Monitor Mode, and making active commands as Boss Mode. We then conducted follow-up semistructured interviews to reflect on their experiences with these modes, gain a deeper understanding of their preferences, and brainstorm ideas for robot control and interaction.

The results in this paper capture desirable user controls and interaction patterns with an autonomous robot. Findings from the navigation tasks revealed that blind participants often exercised Monitor Mode to assess the environment in crowded situations, followed by Boss Mode to play an active role in engaging with crowds based on information provided by the robot. The interaction between participants and the robot also involved negotiating movements, as participants attempted to understand the robot's capabilities and irregular movements. Follow-up interviews further revealed that participants were willing to take control of their interactions with bystanders by intervening in the robot's navigation. They expressed their desire to navigate more complex situations through mutual support and communication in user-robot interactions, such as the robot frequently explaining its movements, requesting human assistance, or issuing alerts on behalf of users when needed. The robot's role was discussed to be both supportive and collaborative.

The main contribution of this work is empirical. Through direct observation and reflection, we gained insights into user controls, commands, and interaction strategies from the perspective of blind participants using the autonomous navigation robot in real-world environments. Participants actively negotiated movements with the robot, switching between control modes like Monitor Mode to assess surroundings and Boss Mode to take charge of crowd interactions. This study highlights opportunities to develop interaction modes that allow users to retain agency in social navigation, particularly in complex environments, while still benefiting from the robot's capabilities. These insights can be used to shape future social robot designs that foster collaboration and trust between users and autonomous systems.

1 Related Work

There is a rich literature on autonomous navigation robots to support blind people, primarily equipped with capabilities for destination navigation [2, 19, 64, 89], obstacle avoidance [39, 58, 78], a combination of both functions [16, 33, 73, 79, 82], and environmental recognition [49, 82]. The underlying promise is that blind users can benefit from this assistance where they only need to follow the robot's movements [33] both outdoors [16, 58, 73, 82] and in indoor spaces [39, 47, 49, 79]. However, these movements are often unpredictable for blind users when following [33] and may heighten feelings of *alienation* if they cannot influence the interaction and outcomes [28]. In the context of broader navigation research, active user engagement has been discussed for its benefits [21, 22, 48].

In this section, we contextualize our work within the broader field of Human-Robot Interaction (HRI) as well as prior efforts in accessibility. Within the context of accessibility (Sec. 1.1), we explore previous work that deviates from the typical robot-following approach and focuses on advancing control and autonomy. We situate these efforts within human-centered frameworks [32, 71] to discuss related design metaphors that can be applied in autonomous robot navigation for blind people. We then discuss literature on social robot navigation (Sec. 1.2), bringing in unique challenges for the exploration of user control. In crowded environments with pedestrian traffic and queues, a robot's behavior is determined by a balancing act between safety and social compliance. Such settings demand more opportunities for control and autonomy.

1.1 Control in Human-Centered AI & Robotics

The design of HRI and the integration of AI within these systems have been explored from various perspectives, with a growing emphasis on human-centered frameworks [68, 71]. Designers, developers, and other stakeholders are moving from the idea of autonomous agents to a more nuanced approach that considers human agency and collaboration [17, 32]. Ben Shneiderman's framework [71] emphasizes metaphors like "supertools" and "active appliances," positioning AI as a means to amplify, augment, and enhance human performance rather than replace it. This metaphorical approach is central to designing systems that maintain human control even in highly automated environments. While full robot control can have its advantages for blind people, especially in unfamiliar environments [31], the lack of control on technology and its assistance may create dependency on others when it fails [44]. This is particularly concerning given that blind people already have limited access to spatial information, which poses significant barriers to independence [6, 30]. To address these challenges, studies have suggested an interdependence framework that supports collaborative relationships between blind people and their surroundings [40] or assistive technologies [7].

Our research builds on the existing knowledge of human control and design metaphors in human-centered frameworks, applying these concepts to the context of blind users and autonomous navigation robots. Efforts exploring this dimension with blind people (*e.g.*, [64, 89]) have shown that balancing automation with human control is crucial for fostering users' confidence, comfort, and trust. Zhang *et al.* [89] also introduced the unique perspective of *"I am the follower, also the boss"* reflecting blind users' experiences with

different modes of autonomy with robotic companions. A similar dynamic can be seen in the roles of guide dogs and blind people [38]. While guide dogs can be considered *"the eyes that lead"* [77], both parties engage in a collaborative effort for orientation and mobility [83]. For example, blind people are responsible for general orientation—monitoring their surroundings and making navigational decisions by issuing cues (*i.e.*, commands) to their guide dogs [38]. Similarly, blind users interacting with autonomous robots require a degree of control, such as the ability to intervene in decision making or adjust the robot's behavior.

No single mode of autonomy is optimal in every situation [64], making this balance between automation and user control critical. Despite the ongoing efforts to improve control for blind people interacting with autonomous robots [64, 89], identifying interaction mechanisms is still underexplored, especially in the context of social robot navigation. It involves complex settings that demand attention to social norms and the user's sense of agency.

1.2 Social Robot Navigation

There are many challenges in building and deploying robot navigation technologies in human environments. The survey paper by Mavrogiannis *et al.* [57] categorizes the challenges into planning, behavioral, and evaluation challenges. Among the three, planning and behavioral challenges are closer to this work.

From a planning perspective, researchers see these challenges with the need to balance efficiency and safety [27, 57]. When safety is prioritized, planners often encounter the "freezing robot problem" [76], where robots stop unnecessarily even when navigable paths exist. This occurs because safety prioritization leads to inflated obstacle zones around pedestrians, and all viable paths in the robot's planning space are blocked. To address this issue, researchers eacho that human-robot interaction modeling is essential [76] and have leveraged learning-based methods [20, 53], classic model-based approaches [25, 74], or hybrid techniques [80, 87]. Despite the recent advances, the "freezing robot problem" is an ongoing challenge [57].

Our work is situated within the social navigation challenges, including the "freezing robot problem," by exploring the concept of control and investigating how blind users of guide robots can engage with this issue. We also delve into the social behavior aspect of human-robot interaction. Existing research in this domain has studied a wide variety of behaviors, such as pedestrian group dynamics [81], proxemic zones [5, 34], robot escape strategies [15], expression of robot intention [52], and line waiting [61]. However, this remains an area lacking established principled knowledge to guide robot design in social navigation [57].

To address this gap, we broaden our understanding of social behaviors in navigation through a different lens. Numerous studies have examined blind people's experiences in navigating real-world environmentse [10, 14, 84, 85], providing valuable insights to inform the design of technology that can support and enhance these experiences. Understanding their navigation strategies has also guided the development of assistive technologies [18, 41]. In this work, we adopt a similar methodology; as an initial step, we aim to understand blind people's social navigation experiences and strategies, with the goal of applying these insights to the design of social navigation robots with greater user control.

2 Preliminary Interviews

Leveraging a human-centered design approach, we draw on human-human interactions as key references to inform the design of human-robot interactions [71]. We conducted a preliminary study with blind people using structured interviews to understand their social navigation experiences and strategies in everyday situations. Insights were intended to guide the design of our user study involving navigation tasks with an autonomous robot, exploring their choices of control in complex real-world environments. Given this purpose, we framed this preliminary study around two key questions: (i) What challenges do you encounter when navigating around people, such as in crowded environments? and (ii) What strategies do you use to mitigate these challenges? Responses to the first question informed the social navigation scenarios for the tasks, while the second provided insights into interaction strategies for designing social robot navigation (introduced as probes for user commands in Sec. 3.3.1).

2.1 Recruitment and Participants

Our studies were approved by our institution's Institutional Review Board (IRB). To advertise our preliminary study to the local blind community, we used (a) an existing mailing list associated with the museum in Japan—our collaboration institution included in the IRB and (b) visitor trials of the autonomous navigation robot deployed at the museum, where participants were invited for a separate interview. We designed the study to be structured with two focused questions and brief demographic questions, aiming to last no more than 20 minutes, and called for volunteer participation. We obtained their consent prior to audio recording their interview responses.

We recruited a total of 14 blind participants (6 women, 8 men), aged between 32 and 75 years (M = 52.79, SD = 12.15), with 12 participating remotely and 2 in person. Nine participants were totally blind, and five were legally blind. Three participants were blind since birth, and four were blind from a young age with the onset varying between ages 3 and 10. All participants were white cane users, and there was no participation from blind people who use guide dogs. This might be due to the small community in Japan, which is reported to have about 1/10 of the number of guide dogs available in the U.S. [62].

2.2 Findings

We used a deductive thematic analysis approach to examine participants' responses [11]. This method involved reviewing and organizing the responses based on predefined themes that focused on challenges and strategies related to social navigation, as derived from the focused interview questions. Through an analysis for recurring patterns and similarities, we identified two overarching scenarios that reflect major navigation challenges and workarounds to address these challenges in each context.

(1) Streams of People. Participants reported difficulties in navigating crowded spaces, which they described as 'streams of people.'

This complexity was reported in various scenarios, including train stations, riverfront trails, hospitals, busy shopping districts, festivals, and other events. A key issue was their struggle to adapt to the unpredictable behaviors of the crowd, making it hard to tell whether others were walking toward or away from them, stopping, or forming a line. Being swept along by the crowd and feeling physically close to strangers was particularly frightening, often resulting in a loss of direction and orientation. Participants also mentioned that relying on audio cues, such as those from station gates or escalators, becomes less effective in crowded areas.

Maintaining a safe distance was challenging as well, as participants could inadvertently trip others with their white canes or risk damaging the cane. In dense crowds, the limited ability to move the cane from side to side complicates obstacle detection and maintaining a straight path. Although we initially anticipated that this might be different for those who use guide dogs—typically trained to maintain a straight line and avoid obstacles [77], they too can experience challenges when navigating through streams of people. A participant from our main study later reported that crowds often fail to notice guide dogs, leading to safety issues due to the unpredictable behaviors of people around.

Participants shared various workarounds to navigate these situations. One common strategy was **asking others** for assistance. They also used localization techniques, such as **following physical cues** like tactile paving or **stopping temporarily** to gather information. **Following environmental cues**–like sounds from nearby shops or escalators, or even smells and wind–were considered helpful for navigation. Additional strategies included walking along walls or listening for reflective sounds, such as changes in echoes or audio announcements, to avoid collisions and identify crowded spaces. While this awareness of the crowds was intended for caution, participants also reported **following crowds** (*e.g.*, people heading toward an escalator after getting off a train) as a way to find their direction. However, it was often described as a frantic process, where blind people were not intentionally recognizing the crowd but simply moving along with others as they were in close proximity. **Alerting others**, such as by tapping a white cane, was also a technique used to interface with the crowd, though it was not ideal.

(2) Lines. Participants expanded on difficulties with identifying and following people in line. They reported uncertainty about where to stand, how the line was forming, and whether it was moving, especially in more complex queues like U-shaped or forked lines. They also found it difficult to maintain social distance and follow social rules. There was concern about unintentionally cutting in line or getting in the way of others who were already in line. These challenges were noted in various contexts, including waiting for buses, taxis, trains, escalators, ATMs, cash registers, restrooms, hospitals, and food distribution shelters.

Workarounds for standing in line and tracking its movement included **asking others** in line, similar to how they navigate through streams of people. They also asked store staff for information about the end of the line or to notify them when it was their turn. To gauge the line's movement and positioning, participants reported **following audio cues** from footsteps or conversations of people in line. Assistive technology also played a role, including sonar devices and remote sighted services. To help move forward in the line, a white cane was used to gauge the distance between people in front. However, it often led to accidental contact and was considered a last resort when other methods were ineffective. They ultimately ask for sighted assistance, particularly in settings like hospitals or grocery stores, where standing in line is expected and there is a greater need for smooth queuing and navigation.

3 User Study: Navigation Tasks and Semistructured Interviews

This next study is an exploration including navigation tasks and semistructured interviews to elicit blind participants' preferences and needs for control with a navigation robot. Building on the social navigation challenges identified in our preliminary interviews, the navigation tasks focus on two key scenarios: streams of people and lines in real-world environments. These scenarios are particularly challenging for both humans and robots due to the 'freezing robot problem' [76] and potential misalignment with social behaviors. During the tasks, participants explore strategies for navigating these situations, using different modes of autonomy called Omakase, Monitor, and Boss (defined in Sec. 3.3). We also probe participants on strategies they could employ, referring to them as commands such as 'alerting' or 'following,' based on insights from the preliminary interviews. After completing the tasks, participants reflect more broadly on what actions they would take in such scenarios and what support they would expect from the robot to enhance their social navigation strategies and experiences.
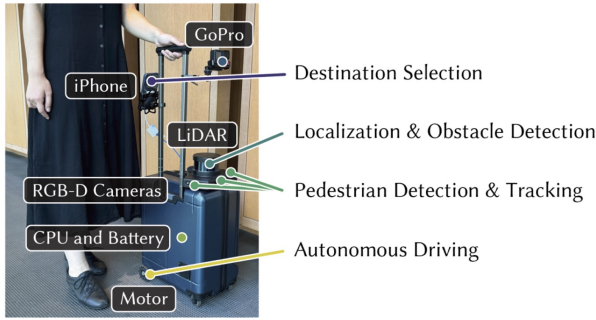
To design our study and enable participants to experience the social navigation scenarios, we conducted our study at a museum where an autonomous robot has been deployed to assist visitors, primarily blind people. We chose the museum environment for its real-world complexity and relevance, with an anticipated average of 2969 visitors per day during the study period, based on statistics from August 2023.

3.1 Recruitment and Participants

We recruited participants through a mailing list of nearly 200 blind or low vision people. The eligibility criteria were as follows: people who regularly navigate with a white cane or guide dog; aged 18 and older; ability to go to the study site (the entrance of the museum) by themselves or with others *e.g.*, sighted guides. They were compensated 7500 yen for their time (2.5 hours) and transportation expenses. After screening the initial responses to the call for participation from 19 subscribers, we invited a total of 13 participants, as listed in Table 1. Among these participants, three experienced the navigation robot for the first time. Among those with experience, the number of times they experienced the robot ranged from one to six. While nine participants have visited the study location in the past (either as personal visits to the museum or other research study visits), no one reported being familiar with the environment to navigate on their own using their mobility aid. P3, 10, and P13 also participated in the preliminary interviews. As participants with other impairments, P8 had dwarfism and slight difficulty in bending their joints; and P12 had mild hearing loss and used a hearing aid.

Table 1: Demographic information of user study participants.

ID	Age	Gender	Vision Level	Impairment Duration	Mobility Aid(s) (*: primary)	Solo Travel Frequency
P1	73	man	legally blind; light perception	>10 years	white cane	few times/week
P2	49	woman	totally blind	>10 years	white cane	5 times/week
P3	54	man	totally blind	>10 years	white cane	everyday
P4	n/a	woman	low vision; no peripheral vision	>10 years	guide dog* & white cane	5 times/week
P5	51	woman	legally blind; hand motion	5 years	white cane	4 times/week
P6	60	woman	totally blind	>10 years	white cane	few times/month
P7	n/a	woman	totally blind	>10 years	white cane	5 times/week
P8	58	man	legally blind; light perception	>10 years	white cane	2–3 times/week
P9	44	woman	totally blind (left); no peripheral and color vision (right)	>10 years	white cane	everyday
P10	58	man	totally blind	>10 years	white cane	everyday
P11	59	woman	totally blind (right); some light perception (left)	>10 years	guide dog* & cane	everyday
P12	42	man	low vision; no peripheral vision	3 years	white cane	5 times/week
P13	60	woman	legally blind; no peripheral vision	>10 years	white cane	everyday

a) Robot's Navigation Functionalities

GoPro
iPhone
LiDAR
RGB-D Cameras
CPU and Battery
Motor

Destination Selection
Localization & Obstacle Detection
Pedestrian Detection & Tracking
Autonomous Driving

b) User Study Setup

Experimenter
Participant
Navigation Robot

Figure 1: a) The robot's key functionalities in navigation and b) the user study setup with the robot, participant, and experimenter

3.2 Apparatus

We used an autonomous navigation robot deployed in the real world for our exploration. The robot's hardware and software are based on an open-source project [24]. Computational resources, including a CPU and battery, are housed in a commercially available suitcase, with motorized rear wheels (Figure 1a). The robot localizes its position and orientation by matching point clouds from its LiDAR sensor to a pre-built map. Once the robot's destination is set via a smartphone, the navigation planner calculates the least-cost path to reach it. During navigation, the robot detects surrounding obstacles using the LiDAR sensor and plans a safe path to avoid them. Additionally, the robot tracks nearby people with three RGB-Depth cameras powered by the YOLOv4 image recognition engine [9], ensuring it maintains an appropriate distance from individuals. Upon reaching the destination, an announcement is made through the smartphone. To record the environment during the experiment, we mounted a GoPro camera on the suitcase handle.

3.3 Study Design

3.3.1 Prompts and Probes. Before starting the navigation tasks, we prompted participants with three modes of autonomy and asked them to use a think-aloud method to communicate their choices of modes during the tasks. This approach was designed to reflect their desired levels of engagement and decision-making. We built on related design metaphors from prior literature, specifically drawing from more culturally aligned perspectives from Zhang *et al.* [89]'s follower and boss roles for blind people interacting with robots. These roles can be further understood through the social robot design dimension [3], which includes a leadership model that determines "Who initiates the interaction? Who determines what happens next?" This model captures a spectrum from robot-led to user-led, with a mutual role reflecting dynamics similar to the relationship between guide dogs and blind people. In this mutual effort, blind people monitor their surroundings to make navigational decisions [38]. The following explains three distinct modes of autonomy, based on these roles, for interacting with the environment and the robot: Omakase Mode, Monitor Mode, and Boss Mode.
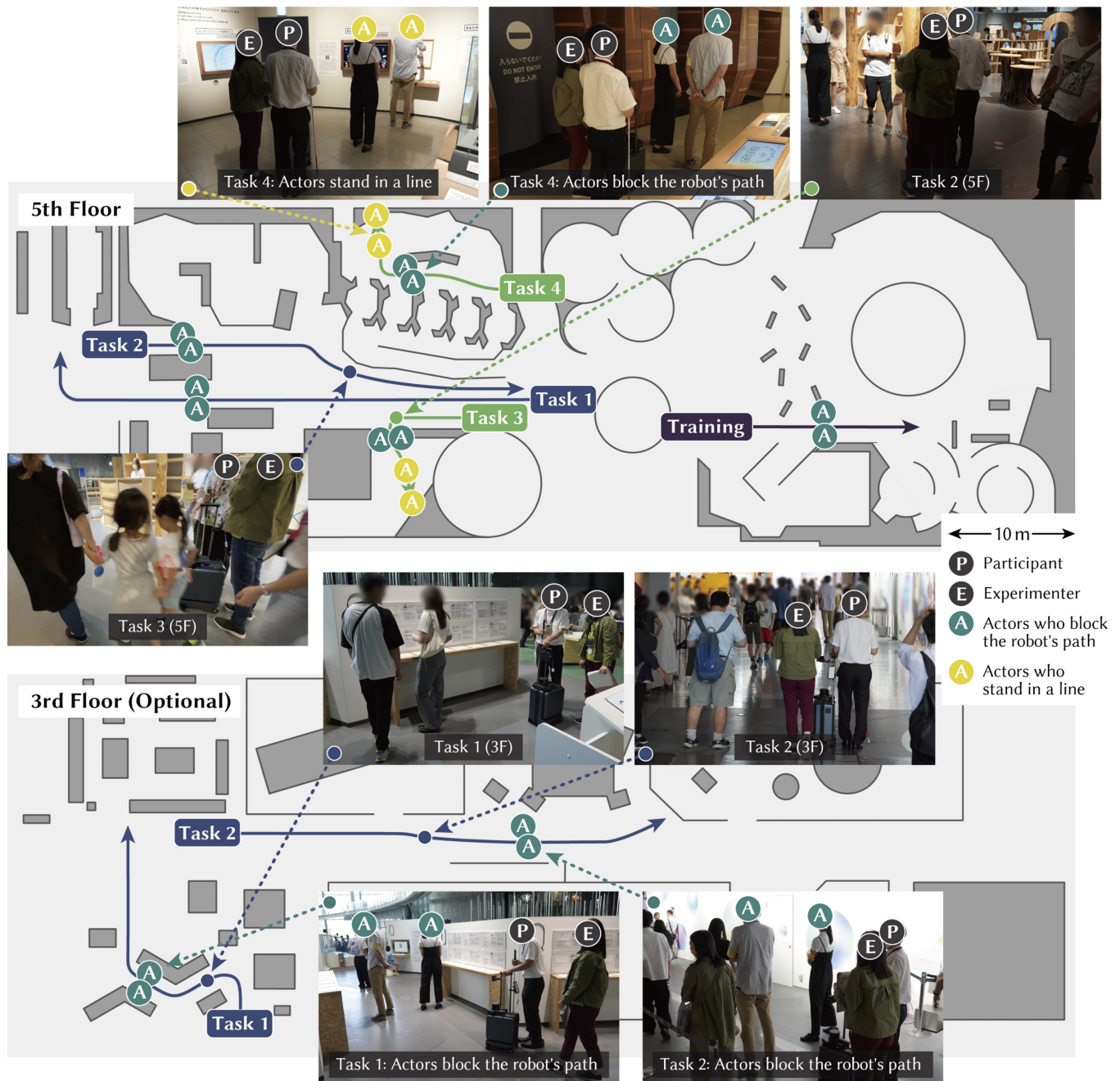
Figure 2: Routes used in the navigation tasks, where participants navigated a crowded environment in the museum with the robot and experimenter. Actors either blocked the robot's path or stood in a line at the destination.

- **Omakase** - Characterized by a low level of autonomy, this mode involves a passive role with minimal decision-making by the participant. For example, this mode is selected when the participant follows the robot's decisions without actively engaging in the navigation process, even if the robot stops due to the freezing problem.

- **Monitor** - Characterized by a developing level of autonomy, this mode involves the participant engaging with their environment by asking questions or seeking information from the robot. For example, this mode is selected when the participant asks questions such as "Are there people blocking the way?" or "How many people are there?" to guide their navigation choices. In this study, this mode is realized using

the Wizard-of-Oz method, where the experimenter responds to user queries based on a predefined set of information the robot can detect, such as pedestrians and obstacles [24]. We adopted this method to encourage participants to reflect on the interaction aspects of this level of autonomy, rather than focusing on the technical aspect. AI-driven conversational models have been explored to assist blind people with scene understanding, calling for human-like Q&A systems [8, 29]. However, users often find responses unreliable due to missing or incorrect information, leading to reduced trust [23, 45]. Introducing such systems is logistically challenging, and addressing these technical limitations is beyond the focus of our study.

- **Boss** - Characterized by a high level of autonomy, this mode involves the participant taking full control of their navigation and employing strategies. This mode is annotated when participants issue commands such as "Alert people" or "Move forward" and interact directly with bystanders by saying "Excuse me" to navigate through crowded areas. The participant's role in directing or choosing these actions is central to this mode.

Additionally, before the navigation tasks, we probed examples of commands in Boss Mode based on our preliminary interviews that revealed blind people's strategies in social navigation. Participants were encouraged to brainstorm their own commands, with the provided list serving only as a guide.

- **User Inquiry/Alert** - Participants ask or alert specific people to move aside or to make their presence known to those around them, such as saying "Excuse me. I'm coming through."
- **Robot Inquiry/Alert** - Participants instruct the robot (which in this study was the experimenter) to ask or alert specific people to move aside, regardless of the method, such as beeping or saying "Excuse me."
- **Movement** - Participants direct the robot's movements, such as waiting and moving forward, which often conflict with the robot's predefined movement rules.
- **People Following** - Participants instruct the robot to follow specific people.
- **Physical Cue Following** - Participants instruct the robot to follow physical features such as a wall or tactile paving.

3.3.2 Procedure. We first explained the purpose of the study, obtained informed consent, and conducted a training session prior to the user tasks. During the training, participants first walked with the robot without taking any actions, experiencing Omakase Mode, and adjusted the robot's walking speed. Then, to explain what participants can do in the other two modes, we created a situation where the robot couldn't move because its path was blocked by two actors. In this situation, they were instructed to ask or negotiate with the robot to understand their surroundings (Monitor Mode) and to make a command to clear the path (Boss Mode).

The robot used in this study can navigate fully autonomously to its destination while avoiding pedestrians and obstacles along the way. It can also wait in line autonomously near the destination. However, in crowded scenarios, the safety-priority design often

causes the robot to stop, which often triggers users to switch modes as discussed in Sec. 3.3.1.

Navigation Tasks. Participants were asked to walk along four specified routes during regular business hours, freely switching between three modes until they arrived at each destination. Figure 2 shows the robot's regular paths (task 1 through task 4), which participants followed in numerical order. Of the four routes, two (task 1 and task 2) were routes to experience stream situations and the other two (task 3 and task 4) were routes to experience queuing situations. Although the user tasks were primarily conducted on the fifth floor of the museum, task 1 and task 2 were performed on the third floor when the fifth floor was not crowded. On each route, two actors completely blocked the robot's path at one point, and for the routes of task 3 and task 4, they also created a line in front of the destination. To artificially create situations where the robot can move in a line, we selected exhibits as the destinations for task 3 and task 4, where the robot is prohibited from calculating a detour due to the narrowness in front. During the tasks, the experimenter walked one step behind the robot and interacted with participants and surroundings on behalf of the robot in Monitor and Boss Modes (Figure 1b). The experimenter responded to participants' questions in Monitor Mode based on predetermined rules for explaining the surroundings, such as "People (or obstacles) are blocking the path (or have just crossed the path)." After completing task 2 or task 4, participants took brief interviews to facilitate post-session interviews. They were asked about any difficult situations they faced during tasks, what they could do in such situations, and what kind of support they would like from the robot.

Post-session Interviews. After completing all of their tasks, participants took post-session semistructured interviews. First, they were asked open-ended questions about how they would resolve challenging situations, including both those they mentioned in brief interviews during the tasks and situations they hadn't experienced, such as meandering lines or single-file queues, which were prepared in advance based on preliminary interviews. They freely bounced around ideas about what they wanted to do, such as possible commands they could issue and what kind of support they needed, including specific information, in each situation. Second, they were also asked to answer a set of questions on a seven-point scale with reasons, which is reported in Figure 6 and Figure 7. Q1–Q3 were Likert-scale questionnaires about each mode (1: strongly disagree, 4: neutral, 7: strongly agree), and Q4–Q8 asked how they preferred the leadership for each command (1: fully robot-led, 4: neutral, 7: fully user-led). Lastly, they were asked about their demographic information, detailed in Sec. 3.1.

3.4 Analysis

Our data analysis was structured around two main phases of the study: user navigation tasks and post-session interviews. We transcribed audio recordings from both phases and used audio and video recordings from the navigation tasks to capture user behavior, interactions, and crowd dynamics.

3.4.1 User Tasks Annotations. We employed a multi-faceted approach to analyze user tasks, annotating the modes of autonomy and issued commands while mapping them to our prompts and probes listed in Sec. 3.3.1. We also annotated user queries in Monitor

Mode and crowd situations during the interaction events, as described below. The coding scheme was refined iteratively within the team. One researcher took the lead on annotations, while at least one other researcher conducted a detailed pass. Any annotation uncertainties were resolved in team meetings or offline discussions.

Monitor queries. We delved into Monitor Mode where users query the robot for information. Our goal is to find out what types of information the users seek in Monitor Mode. We employed a descriptive and qualitative analysis coding procedure to develop annotations. The procedure is a manual iterative clustering-based procedure over two steps. In the first step, we read through the participants' queries and generated concepts (*a.k.a.* information request types). In the second step, we assigned the concepts back to each query. If a query could not be clustered into existing concepts appropriately, we created a new concept and reexamined all previous queries with the updated set of concepts. We repeated the two-step process until all queries could be mutually exclusively clustered into these concepts. Finally, we made sure that a consensus was reached on the final set of concepts and their assignments to the queries. These concepts are our annotations.

- **Obstacles** - Queries about something that participants sensed was blocking the robot's path.
- **People** - In-depth queries about the status of the pedestrians around the robot.
- **Action** - Queries related to the robot's current action.
- **Attempt** - Queries on whether the robot can try something.
- **Situation** - General queries about what is happening or if there are any problems.
- **Line** - Queries related to line formations.
- **Destination** - Queries related to the navigation destination.

Crowd situation descriptions. We grounded our coding of crowd situations on established social navigation literature. In two recent survey papers [27, 57], the most prevalent categorization to divide social navigation scenarios is whether the crowds are stationary or dynamic. It is commonly accepted that static crowds behave distinctively differently from dynamic crowds, and separate rules need to be considered for social navigation algorithms. Additionally, we noted the need to separate line-related scenarios from static crowd situations as a special case. This is also based on established literature [50, 61], which ascertains that crowd behavior in lines is unique and is difficult to tackle by a generic social navigation algorithm. In summary, we coded crowd situation descriptions according to the following:

- **Stationary** - All pedestrians in the robot's close proximity were stationary (*e.g.*, holding conversations, watching exhibits, looking at phones).
- **Dynamic** - All pedestrians in the robot's close proximity were walking in any direction. This also implies that soon after the subjects had initiated the interaction, the pedestrians were likely no longer around.
- **Line** - The subject and the robot arrived at the start of the line or were moving in the line.
- **Other** - Occasionally, interactions were initiated when no humans were around. This typically happened when the robot failed to plan its course promptly.

Note that there were rare scenarios where both static and dynamic humans were present. In this case, the responses gathered were used in the analysis of both static and dynamic cases, because we believe the participants' behavior was influenced partially by both types of pedestrians.

3.4.2 Post-Session Interviews. The interviews included both open-ended questions for design discussions and Likert scale questions with accompanying shorter qualitative responses.

For the qualitative analysis of the open-ended discussions, we employed inductive thematic analysis [11]. The process was led by researcher R1 who facilitated all discussion sessions. R1 began by assigning descriptive (*i.e.*, semantic) codes to the transcripts of participant comments and feedback. These codes were then organized into categories. R1 proceeded to generate initial themes and sub-themes based on these codes, integrating interpretations throughout. In parallel, researcher R2 conducted the analysis of Likert scale responses. R2 looked into the spread and variability of responses using medians to represent central tendencies and interquartile ranges (IQR). R2 also summarized how participants selected each Likert scale option and identified general trends in their comments.

The preliminary findings from both analyses were refined through team meetings for a more unified understanding of the concepts in the data [12]. R1 and R2 complemented each other's analyses; for example, R1 identified an initial theme related to *user involvement* in human-robot interactions, which was broadened with R2's summary of responses concerning independence and control. In the following section, we present our findings from the interviews, focusing on participants' perspectives on autonomy and shared control. These insights are contextualized by the user tasks they performed with the autonomous robot prior to the interviews, which helped them reflect on their interactions and discuss future interfaces.

4 Results

4.1 Choices for Autonomy and User Queries & Commands in Navigation Tasks

In this section, we report our observations of how participants explored and transitioned between different interaction modes for autonomy to complete navigation tasks in various crowd situations (Sec. 4.1.1), as well as their queries (Sec. 4.1.2) and commands (Sec. 4.1.3) to exercise more control outside Omakase Mode.

4.1.1 Crowd Situation and Mode Transition Analysis. Figure 3a shows the ratio of transitions to Monitor or Boss Mode from Omakase Mode. We observed that participants often transitioned to Monitor Mode from Omakase Mode, such as asking for information about their surroundings or the reason the robot stopped. These situations involved the robot's path being blocked (Stationary: 89%), a few passersby walking across the robot's path (Dynamic: 76%), and people forming lines (Line: 93%). Other situations (62%) also led participants to switch to Monitor Mode, partly due to feedback and interface disruptions that made it difficult to understand why the robot stopped, such as audio feedback delays for indicating destination arrival or the deactivation of the touch sensor on the handle when participants were not consciously gripping it.
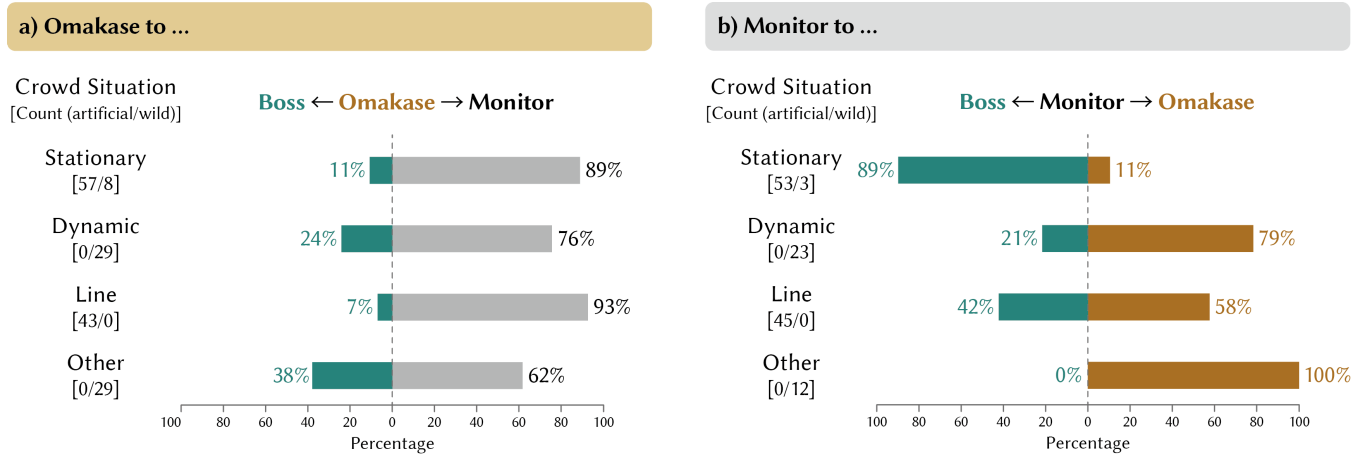
Figure 3: Percentage of transitions between modes of autonomy by crowd situations: a) from Omakase Mode to Boss or Monitor Mode and b) from Monitor Mode to Boss or Omakase Mode.

When participants transitioned to Boss Mode directly from Omakase Mode, we observed two main types of situations. First, participants, having already understood the environment as crowded or detected people nearby based on surrounding sounds, issued commands directly to address the robot's movements, such as it zigzagged unpredictably or stopped for a long time. This was particularly common in stationary or line situations. Second, some participants (notably P10) constantly gave verbal alerts (*e.g.*, "Excuse me") to their surroundings while walking, regardless of the situation. This led to an increased frequency of transitioning into Boss Mode in dynamic and other situations.

Figure 3b shows the ratio of transitions to Omakase or Boss Mode from Monitor Mode, indicating participants' choice of autonomy after gaining spatial information. Participants tended to shift to Boss Mode in stationary situations (89%) and to Omakase Mode in dynamic and other situations (79% and 100%, respectively). In stationary situations, the trend of choosing Boss Mode was not surprising, as the robot's path was blocked (often by the actors in our study), leaving participants with no option but to ask them to step aside. In dynamic and other situations, Omakase Mode was optimal, as the robot's path was eventually cleared with no further action needed. For example, P7 asked, *"What's wrong?"* to which the experimenter responded, *"The robot tried to take a detour because people had just crossed the path."* P7 shifted to Omakase Mode after assessing the situation.

In line situations, the ratio of transitions to Omakase Mode from Monitor Mode was slightly higher than that to Boss Mode. This was because after assessing the line situation upon reaching the end of the line, participants chose to follow the robot's movement without issuing any commands. Yet, there were instances where participants gave commands to the robot to follow the person in line. Additionally, when the line moved, some participants queried the current status of the line, such as the number of people waiting, but typically returned to Omakase Mode without giving further commands. For example, when P12 arrived at the end of the line, they asked, *"Is this a line, or are people just standing in front?"* and the experimenter responded, *"There are people standing in front of*

us, these are people waiting in line." P12 followed up with *"How many people are in line?"* to which the experimenter responded, "Now there are two people in front of us in line." P12 then confirmed, *"OK, I will shift to Omakase Mode."*

4.1.2 User Query Analysis. We analyzed what types of queries participants made under different crowd situations in Monitor Mode. We observed varying distributions of information request types conditioned on each of the four crowd situations during Monitor Mode, as shown in Figure 4.

In stationary crowd scenarios, participants mostly queried about obstacles (22%) and tried to explore what the robot could attempt in such situations (23%). When making obstacle-type queries, participants often sought confirmation, such as *"Is there something in front?"* In other cases, participants inferred that the robot had stopped due to obstacles and asked for more details, such as *"What is in front?"* Regarding attempt-type queries, when stopped, participants asked if the robot could try taking a detour to the left or right, or if it could proceed forward through the blocking obstacles or people. This might indicate that participants used Monitor Mode to explore the robot's capabilities and negotiate strategies for navigating the obstacles. Another inquiry type relevant to the stationary crowd situation was about the line (14%). In waiting-in-line tasks, participants were unsure if the robot had stopped because of the line, prompting questions such as *"Is this a line?"* or *"Are people waiting in line or just standing around?"*

In dynamic situations, participants made distinctively different types of queries. Action-type queries were dominant in these situations (24%) compared to stationary situations (13%). This happened often due to the robot taking zigzagging paths when navigating in dynamic crowds, which prompted the participants to question its actions, such as *"Why are you turning right?"* or ''*Why can't we go straight?"* Interestingly, we see a higher percentage of people-type queries (20%) compared to those made in stationary situations (13%). They could sense when pedestrians were no longer present nearby and asked to confirm with questions such as *"Did someone just pass by?"* On a related note, in dynamic situations, obstacle-type
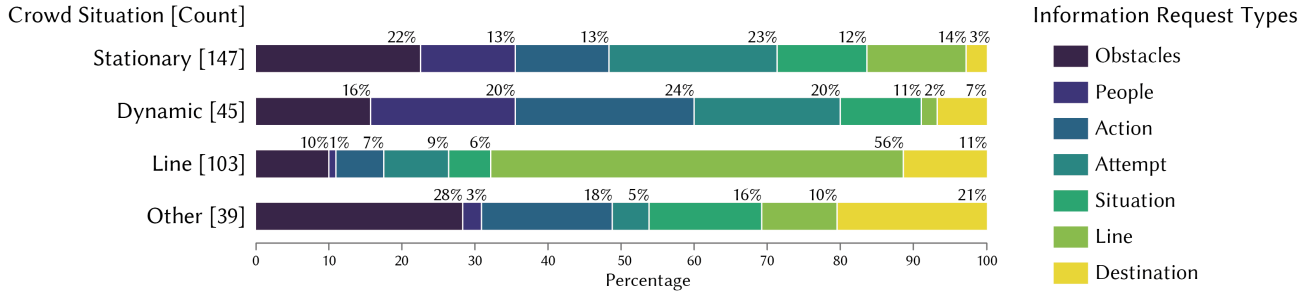
Crowd Situation [Count]

Information Request Types

Figure 4: Percentage of information requests by crowd situations in Monitor Mode.

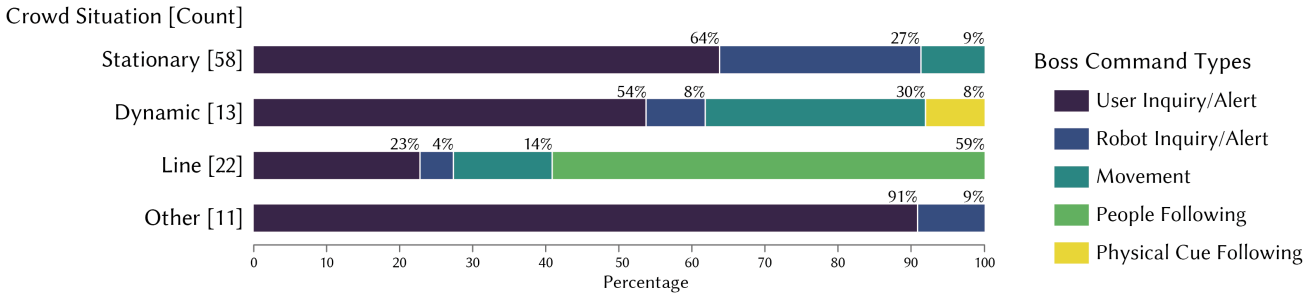Crowd Situation [Count]

Boss Command Types

Figure 5: Percentage of user commands by crowd situations in Boss Mode.

queries were less common (16%), as participants often assumed the obstruction was caused by people in the surroundings. Therefore, they mostly inquired for more details about people, asking *"Is it crowded?"* or *"How many people are there?"*

Unsurprisingly, line-type queries dominated the line crowd situations (56%). Upon closer inspection, most participants wanted to confirm whether they were in the line (35% of line-type queries). Other common queries included *"How many people are in the line?"* (17%), *"Are we at the start of the line?"* (12%), *"Are we at the end of the line?"* (10%), and *"Let me know when the line moves"* (9%). At times, participants didn't query and directly issued commands to the robot to follow the line.

"Other" crowd situations typically took place due to occasional technical failures or the robot taking time to chart its course, which often confused participants. This confusion is reflected in the increased amount of obstacle-type queries (*e.g., "What is in front?"* 28%), action-type queries (*e.g., "Why did you stop?"* 18%), situation-type queries (*e.g., "What is the situation?"* 16%), and destination-type queries (*"Have we arrived?"* 21%).

4.1.3 User Commands Analysis. We analyzed the commands participants issued in different crowd situations during Boss Mode. Figure 5 shows the ratio of each command type across these situations.

Inquiry/alert commands were often made in stationary and dynamic situations (91% and 62%, respectively), with user inquiry/alert commands being particularly preferred in both situations (64% and 54%, respectively). This meant participants actively engaged with passersby to clear the way, such as by using direct alerts *"I'm coming through"* or *"Please move to the side"*, or by asking more politely

"Could you move out of the way please?" Robot inquiry/alert commands involved speaking similar phrases on behalf of users and making a beep sound. In dynamic situations, movement commands were also common (30%). These commands typically involved asking the robot to move in a specific direction, wait until the crowd cleared, or return to a point where it had been moving smoothly and reroute.

In line situations, people-following commands were the most frequently used (59%). Robot-led movement commands related to standing in line were also made (14%). User inquiry/alert and robot inquiry/alert commands happened when participants mistakenly instructed people to move aside, not realizing they were actually in line.

Other notable commands included physical cue-following commands, which we observed by P11 issuing the instruction to *"walk along the edge of the walkway"* to feel relieved by touching the wall. P10 frequently made verbal alerts regardless of surroundings, resulting in the ratio of user inquiry/alert commands accounting for 91% in "other" situations.

4.2 Perspectives on Autonomy and Interfaces for Control

In this section, we reflect on participants' discussions on autonomy and interaction design for greater control in autonomous robot navigation. We also report together the trends in their autonomy choices and interaction preferences based on the questionnaire results. Figure 6 and Figure 7 summarize the results from the Likert scale questions on participants' preferences for each mode (Figure 6, Q1–Q3) and preferences for using boss commands in either a robot-led or user-led manner (Figure 7, Q4–Q8). Although 9 out of
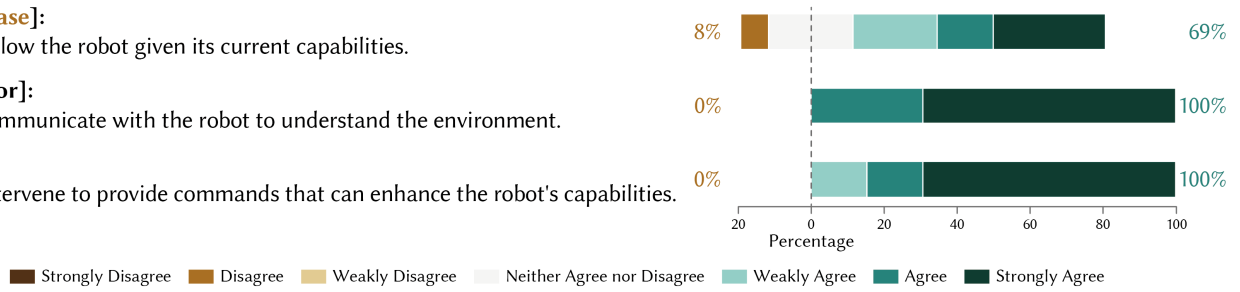
Q1 [Omakase]:
I want to follow the robot given its current capabilities.

Q2 [Monitor]:
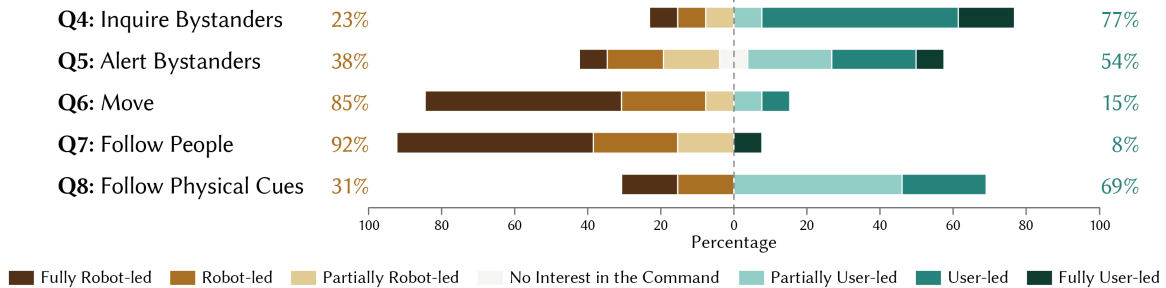I want to communicate with the robot to understand the environment.

Q3 [Boss]:
I want to intervene to provide commands that can enhance the robot's capabilities.

8% — 69%
0% — 100%
0% — 100%

Strongly Disagree · Disagree · Weakly Disagree · Neither Agree nor Disagree · Weakly Agree · Agree · Strongly Agree

Figure 6: Overall experience ratings toward different modes of autonomy.

How would you like to use the following commands—robot-led or user-led?

Q4: Inquire Bystanders — 23% — 77%
Q5: Alert Bystanders — 38% — 54%
Q6: Move — 85% — 15%
Q7: Follow People — 92% — 8%
Q8: Follow Physical Cues — 31% — 69%

Fully Robot-led · Robot-led · Partially Robot-led · No Interest in the Command · Partially User-led · User-led · Fully User-led

Figure 7: Participants' preference for commands on a scale from fully robot-led to fully user-led, with a midpoint indicating no interest in the command.

13 participants agreed that they want to follow the robot (Omakase Mode) given its current capabilities (Q1), all participants expressed a preference for using Monitor Mode to understand the surrounding environment (Q2) and Boss Mode to enhance the robot's capabilities (Q3). This section explores in detail how participants envision interacting with these modes.

4.2.1 User vs Robot Involvement Beyond Omakase. One recurring theme in the interviews was the active role blind participants sought to play in social navigation, rather than being entirely dependent on the robot. A key factor steering this perception was their sense of agency—having control over their actions and outcomes [60]. For example, P13 said, *"As long as we understand the situation, humans should make the decisions. I think it's better for humans to make the decisions, rather than leaving everything to robots, because after all, people should be the ones in control."* As a result, all participants agreed that they want to intervene to provide commands that can enhance the robot's capabilities (Q3 in Figure 6). Four participants (P4, P11–P13) mentioned that the reason was because they wanted to take an active role in the decision-making process, rather than leaving everything to the robot. Regarding the command-related questions (Q4–Q8, Figure 7), participants who preferred user-led control referred to user autonomy as the reason for their preference.

Participants recognized that some complex social navigation situations require more user involvement, and direct communication with bystanders is crucial. This includes scenarios like standing in a multi-line queue or interacting with a crowd with specific needs. For example, P12 mentioned navigating shared spaces with individuals using strollers, canes, or wheelchairs, where direct communication is necessary to negotiate a clear path.: *"It can be hard to ask them to move aside...It's about having a bit of communication where both parties adjust slightly to create a path for each other."* Opting for direct communication also reflected skepticism about whether the robot could fully understand social complexities or communicate effectively on behalf of users. P4 said: *"Because even if you let the robot make the decision, I wonder if it would really understand... For example, like the line we mentioned earlier, if you ask the robot, 'Is this the line for such and such?' and it says, 'No,' you move on, and if it says, 'Yes,' you still can't be sure if it's really a line or not. In the end, I think it's better to just ask the people around you directly."* Therefore, human intervention was considered a more efficient and reliable strategy in social navigation. A similar trend was observed in responses to questions regarding commands for inquiring/alerting bystanders (Q4 and Q5 in Figure 7). For Q4 (inquire bystanders), the median response was 6, with 10 out of 13 participants preferring user-led control (rating 5 or higher). For Q5 (alert bystanders), the median response was 5, with seven participants favoring user-led control. Five participants (P2, P7–P10) explained that their preference was due to the user's ability to handle complex situations more smoothly than the robot.

Additionally, participants' concerns about social acceptance contributed to their preference for human interaction. It was more appropriate for them to engage with bystanders, as opposed to relying on the robot to interface with the crowd such as through alerts. One participant (P10) explained this sentiment: *"I think the robot shouldn't issue warnings unless there's a sense of urgency. I don't feel*

I have the right to assert myself that much." This highlights the perception that using the robot for alerts or crowd interactions might come across as assertive in everyday social contexts. This is also reflected in participant responses to Q5 (alert bystanders), where the majority preferred user-led control, with four participants (P2, P3, P6, P11) expressing concerns about social acceptance.

Nevertheless, participants described robot involvement in navigation tasks as effective in certain situational contexts. For example, P13 appreciated the robot's navigation technology: *"Sometimes, I just don't feel like asking for help. That's why having a robot that can move forward on its own would be amazing."* P3 also highlighted situations where user communication might fail: *"If people don't hear me or are engaged in other activities and don't move aside, and considering my limited ability to speak foreign languages, I thought it might be helpful if the robot could use warning sounds or announcements like 'I am coming through' or 'Please move aside.'"* Others like P4, P5, P10, and P11 echoed this sentiment, emphasizing that in crowded environments, robots generating non-intrusive alert sounds can effectively draw attention and facilitate smoother passage, as P5 stated, *"it would be less embarrassing if the robot played music or said something like 'passing through' instead of me calling out myself."* Participants like P1 and P12 further supported this idea, suggesting different alert options depending on the area's crowd density. P12 explained: *"If there are just one or two people, a simple alert to get them to look and move aside would be enough. However, in busier places where many paths overlap, even if an alert goes off, people might not realize it's directed at them. They could just ignore it, thinking it's a random sound like a smartphone going off,"* proposing that the robot should say 'AI suitcase is passing through' in areas like train stations. While non-visual feedback, such as beep sounds, is actively explored for guiding blind people [46, 66], we found that participants were more interested in a mixed approach with both user and robot involvement. As P6 suggested: *"Before I even have to say anything as the user, the robot could automatically make a sound like 'beep beep' to issue a warning. That way, I could also become aware that there's someone there through the sound, and then I could say, 'Excuse me, could you let me through?'"*

4.2.2 Boss Interaction and Robot Support. Participant feedback on the boss interactions in social robot navigation exposed several strategies, including those demonstrated in the navigation tasks under Boss Mode (Sec. 4.1.3). Participants envisioned active user involvement by asking or alerting bystanders themselves (*e.g.*, saying *'Excuse me'* or inquiring *'Where is the end of the line?'*). We also identified their need for support from the robot to complement their interactions with other people, such as locating them in their surroundings (*e.g.*, "In which direction are they?"). For example, P11 shared their experience: *"Figuring out whether it's a person or not is the most crucial thing. Quite often, we end up saying things like 'sorry' to a pole, and we apologize, even to walls, saying 'I'm sorry, excuse me,' no matter what we bump into."*

Additionally, we collected various comments related to the types of information needed about people nearby—often to ask for help if necessary. These included proximity (P4: *"If they're within 50 centimeters, I might ask them directly"*), activity level (P13: *"If they're walking, they might just pass by but if they're standing still, there's a possibility they could respond."*), orientation & attention (P7: *"I can't*

tell if they're facing away or looking toward me, so I'd like to know that."*), and age & ability to communicate (P8: *"Are they in a certain age group? Also, whether or not they can speak <the same language>, perhaps"*). In their navigation tasks under Monitor Mode (Sec. 4.1.2), we similarly observed people-type queries related to status (*"Have they moved or not?"*), proximity (*"How far are they?"*), activities (*"What are they doing?"*), and attention (*"Have they noticed us?"*).

While such information about nearby pedestrians was often intended to facilitate interactions with the crowd, it was also used to avoid direct engagement. Aligning with blind people's information need relating to children [42], P11 explained, *"Children are at a lower height and their movements can be unpredictable. Much like when I'm with guide dogs, the mindset changes depending on whether children are present or not...this awareness can be helpful."* P1 also made a query about children in Monitor Mode during the navigation tasks (*e.g.*, *"Are there children?"*).

Strategies for movement, also observed in the navigation tasks under Boss Mode, were discussed for flexible navigation through crowds, reflecting a unique relationship between the blind user and the robot. This relationship is collaborative: the robot provides essential feedback, while the user maintains control. For example, P8 highlighted that the robot could monitor and report the situation, while the user would decide when to proceed or halt: *"If the obstacle is moving, you could activate the 'wait mode' if the robot is moving left and right, unable to find a clear path. And if the obstacle is stationary, you could 'monitor' it and switch to 'boss'. I think you'll probably be able to choose between those options."* In another example, P6 explained: *"I think it's possible to ask which route is more open (to the robot) and use that information to give instructions for a detour."* Such communication for movement strategies resembles the dynamics of collaboration between guide dog handlers and their companions [38].

Following cues that could be detected by the robot (*e.g.*, movements of other people) or the user (*e.g.*, tactile paving) was another important strategy. This was not only for navigating complex social environments but also for enhancing safety and comfort, as reflected in the use of the "people following" and "physical cue following" commands in Boss Mode (Sec. 4.1.3). They envisioned instructing the robot to follow the last person in line or maintain an appropriate social distance while waiting in line, hoping to overcome the difficulties as described by P13: *"What I always find troublesome is not being able to see the movements of the person in front of me. I move forward, saying 'Excuse me' with my cane...but then I end up bumping into things and apologizing. Sometimes it's such a hassle that I ask the person in front if I can hold onto their arm to help me get through, but doing this every time is tiring."* The "physical cue following" command, such as following tactile paving or wall edges, was also deemed critical for comfort, as it would help participants walk straight in crowded spaces and alert others to move aside with ease.

4.2.3 Monitor Interaction and Mutual Support. A recurring theme centered around dialogue and negotiation in participants' discussions about interacting with robots. This often involved mutual support to address complex social environments, similar to their interactions with sighted guides providing situational updates. For instance, when moving through narrow spaces, P1 described how

they communicate with the guide to adjust positions: *"With a sign like this, you can move through narrow spaces. It would be nice if a robot could do something similar, like saying 'Please walk behind.'"* In another example, P11 reflected on the robot making a strange movement and indicated their need for communication, shifting away from Omakase Mode: *"When you can't see and the robot makes some strange movements, it can make you feel uneasy, like, 'What's happening?'...If the robot said, 'There are a lot of people, so I'll move in a zigzag pattern,' or just gave a simple explanation like that, you'd understand, 'Oh, the robot is moving like this to avoid people, that's why it's zigzagging.' Then, there would be no anxiety about the movement."*

Given these comments, it is not surprising that all participants preferred Monitor Mode to understand their surroundings (Q2 in Figure 6). P7 rated 7 in Q2 provided the following notable comment: *"By being aware of the situation, people can understand things like, 'Oh, that's why the robot was wandering around earlier,' or 'It stopped because someone was standing in the way.' I think everyone wants to know what those who can see are seeing. If we can grasp that, it feels like walking won't just be walking anymore—it'll become something more enjoyable."*

We also found that mutual support between blind users and robots involved users taking the initiative to assist the robot in navigating complex situations. For example, P13 described how the robot could prompt users to seek help from others when it reaches its limited capabilities, suggesting, *"The robot could tell me, 'I don't know what to do from here, so ask the person behind you,' or 'The line splits from here, and I'm not sure where to line up, so ask the person behind you.'"* This would encourage the user to step in, such as asking other people *"Excuse me, does the line split from here?"*. Similarly, P5 highlighted the importance of the robot communicating difficulties in crowded environments, proposing that if the robot said, *"It's a little difficult to pass through right now"* the user might think, *"Oh, I need to help."*

Additionally, participants' views on interaction with the robot revealed a nuanced understanding of broader societal connections, including concerns that excessive reliance on robots might lead to diminished human engagement. P4 said, *"Even if robots can handle many tasks, people might start thinking, 'Since the robot can do everything, I don't need to help.'"*. P4 also expressed worry that this interaction could reinforce potential stereotypes: *"People might start thinking that blind people can't do anything on their own."* To support the effective use of robotic aids, P6 drew a parallel to guide dog training, emphasizing the need to learn when to rely on the robot: *"Just like guide dog users train with their dogs, I think there's a need to develop a skill set for interacting with the robot...Deciding when to speak for yourself and when to rely on the robot is also part of walking ability."* More so, as users become more familiar with the guide robot, their approach to using the technology could evolve. The quote from P8 illustrated this potential: *"As you become more skilled, the way you use the robot could change. The reliance on Omakase Mode will likely decrease, and a combination of Monitor and Boss will become more common. Initially, you might use the robot's automatic features more often due to uncertainty, but as you become more accustomed, you'll rely less on these features, even in scenarios like waiting in line."* This suggests that over time, users might develop a more balanced approach to interacting with the robot, leveraging active user control to enhance human engagement.

5 Discussion

The goal of this paper is to inform the field of HRI about key considerations for designing navigation robots for blind people, focusing on control frameworks for independence [51]. Our work specifically contributes to interdisciplinary discussions on user control in autonomous navigation, which have involved efforts from human-computer interaction and robotics researchers (*e.g.*, [64, 89]). Based on our findings, we discuss design implications for shared control and suggest directions for future work, while acknowledging the limitations of our study.

5.1 Design Implications for Shared Control

5.1.1 Building Support for Interaction Modes of Autonomy. Our study of human-robot interactions, leveraging different modes of autonomy, revealed key insights into shared control. One aspect was blind users receiving *robot support* and negotiating through *mutual support*. In navigation task sessions, blind participants actively engaged with the robot and the crowd, transitioning between Monitor Mode to assess their surroundings and Boss Mode to directly influence interactions with the crowd. Our post-session interviews further highlighted the need for this combination. Participants discussed *robot support* as a means to exercise their control, using the information it provided to better navigate and interact with the crowd. This dynamic reinforced the idea that the robot facilitates, rather than overrides, human decision-making and agency (see Sec. 4.2.1). This perspective surfaces interdependence as a critical framework to describe relationships between blind users and robots [7]. More so, participants extended this framework to *mutual support*, highlighting instances in which they could assist the robot in navigating complex situations (see Sec. 4.2.3). They asked for situational updates from the robot to gain the ability to explore options for addressing social navigation challenges. We could see this dynamic aligning with previous discussions on sighted guides [43] and guide dogs [38] to realize collaborative efforts with blind users. Indeed, participants in our study referred to their interactions with sighted guides to match their interactions with the robot.

To realize such *robot support* and *mutual support* in autonomous navigation for blind users, incorporating scene descriptions and status announcements is a crucial step. Our observations of user queries during navigation tasks (see Sec. 4.1.2), such as questions about obstacles in the robot's path, crowd density, nearby pedestrian activities, queue starting points, and user position within the line, could guide the design of detailed scene descriptions. Participant feedback (see Sec. 4.2.2) further revealed a need for more specific information about surrounding individuals, including their distance, direction, and characteristics *e.g.*, whether they are adults or children. To address these needs, robots should be equipped with advanced perception modules capable of detecting and identifying obstacles, recognizing pedestrians and their attributes (*e.g.*, position, activity, quantity), and detecting lines. Additionally, supporting effective communication and negotiation is critical, with robots responding to user requests and providing accurate feedback. Leveraging Multimodel Large Language Models [63] may assist in achieving this functionality [45].

Robots should also inform users of potential failures, such as crowd navigation issues or delays in route planning. While a simple

system that announces reasons for the robot stopping has been implemented (*e.g.*, "There is a person ahead") [24], this functionality needs improvement to capture the often complex reasons behind robot stops. This involves improving the robot's ability to interpret complex environments and clearly communicate stop or delay reasons in a way blind users can understand to ease their concerns [37]. Addressing these challenges will require further research in both robotic perception and human factors. A possible solution could use velocity values obtained from the robot's local planner [54], the local cost map around the robot [55], and images of the surroundings as inputs to a Vision-and-Language Model [63] to generate natural language explanations.

5.1.2 Balancing User and Robot Involvement. User involvement was found to be another critical aspect of shared control. Participants expressed a strong preference for taking an active role in their navigation, rather than following the robot entirely, as would be the case in Omakase Mode. This preference was shaped by various factors, including a desire to maintain their sense of agency and address complex navigation challenges. Social acceptance was also a key factor, particularly in scenarios where participants felt it was more appropriate to interact directly with the crowd rather than relying on the robot (see Sec. 4.2.1). This finding connects to the literature on social accessibility (*e.g.*, [70]), which notes that blind people often negotiate boundaries to avoid unwanted attention from those around them [86].

Yet, striking a balance between user and robot involvement was deemed critical. During navigation tasks, participants let the robots take the lead role in certain contexts, such as navigating dynamic crowds or following others in line. Interestingly, these scenarios reflect two primary challenges identified in our preliminary study. We observed that Omakase Mode, with its autonomous navigation, addressed real-world difficulties participants commonly faced.

Moving forward, it is important to consider this delicate balance of involvement in autonomous robot navigation, particularly in longitudinal contexts. Over time, users' relationships with the robots and their preferences for control may evolve. This aligns with broader discussions on user complacency with automation [59]. Interestingly, our findings suggest a unique trend: as participants grew accustomed to interacting with the robot, many envisioned scenarios where they would seek more control, leveraging Boss and Monitor Mode. At the same time, participants reflected on societal implications, expressing concerns about over-reliance on robots (see Sec. 4.2.3).

To address these complexities, personalization in shared control is a much-needed direction. For example, users could choose who leads the interaction—whether the robot or themselves—based on context or preferences. While external communication with surrounding pedestrians, such as non-visual alerts like beeping sounds, have been explored for blind users [46, 66] or other groups [88], the use of adaptive strategies remains underexplored. Kayukawa *et al.* [46] report that blind users would not use sound-emitting systems in environments where silence is expected, such as hospitals and libraries. Our study offers insights into various commands, including alert or inquiry approaches, that take situational factors into account. For instance, participants favored robots taking charge

in noisy, unfamiliar, or urgent situations, reinforcing the need for adaptive designs in shared control systems.

5.2 Limitations and Future Work

Expanding Real-World Scenarios. In this paper, based on the findings from the preliminary study, we prepared two challenging social navigation scenarios with actors and visitors in a museum during business hours. These scenarios were designed to explore the desirable mode of autonomy which varies by situations [64]. During the user tasks, we observed four main types of crowd situations, along with more complex scenarios involving combinations of these types. For instance, there were mixes of stationary and dynamic elements, such as pedestrians walking through groups of standing individuals, or combinations of artificial and natural scenarios, like regular visitors moving aside upon hearing alerts directed at actors. We expect more complex situations to occur in the real world, such as continuous pedestrian streams or coexistence of multiple lines. In these contexts, users' strategies may shift. For future work, we plan to extend our studies to more challenging environments outside the museum, such as train stations and busy shopping districts, which were identified as particularly difficult for blind people in our preliminary interviews.

Revisiting Monitor Mode. In this study, the Wizard-of-Oz approach was used during Monitor Mode, with a human experimenter simulating the robot's responses based on a predefined set of information the robot could detect, such as surrounding pedestrians and obstacles [24]. This method was adopted to encourage participants to reflect on interaction aspects of this level of autonomy, rather than focusing on technical aspects. While recent advances have brought speech synthesis to levels comparable to the human voice [69], limitations in reliability and user trust remain due to potential inaccuracies in responses [23, 45]. Interestingly, in our study, participants were still inclined to issue user inquiries and alerts to nearby bystanders themselves, even though they knew the robot's voice was provided by the human experimenter. This suggests that interaction behaviors in Monitor Mode may not solely depend on the quality or content of the response but on the perceived source of assistance. Future research is needed to understand the interaction strategies and control preferences when assistance is provided by a machine [45]. The perception of assistance as "human-driven" versus "machine-driven" may influence user trust and engagement.

Exploring Long-Term Interactions and User Contexts. Although participants were given a training session to familiarize themselves with the robot's basic interaction (as detailed in Sec. 3.3), some unexpectedly switched to Monitor Mode or appeared confused by the robot's temporary stops in Omakase Mode, even when there were no technical issues or unexpected external factors affecting their movement. This suggests that blind users' interactions with the robot can change based on their familiarity with its functionality and their experience level. In our study, 3 participants were completely new to robot interactions, and 2 had more exposure than others (having experienced 5-6 trials previously with the robot). For those who were using the robot for the first time, the short-term nature of these experiments only reflected initial impressions and immediate responses to first interactions. While some insights into user behavior in the longitudianl context were discussed in Sec. 4.2.3,

further long-term studies involving continued use are necessary to explore how preferences and interaction strategies for user control evolve over time.

On a similar note, with only two participants (P4 and P11) as guide dog users, a more representative user group might reveal different interaction preferences. Future studies should focus on participants with diverse attributes, including those accustomed to walking with robots or guide dog users, to better explore unique interactions, strategies, and challenges.

6 Conclusion

In this study, we explored how user autonomy and interface design in future navigation robots for blind individuals should evolve beyond *omakase*, the current fully robot-led, approaches. Through interviews with 14 blind participants, we identified two key social navigation challenges: navigating through streams of people and following lines, along with the strategies used to address them. We then conducted a user study where 13 blind participants navigated a museum with an autonomous robot, exploring their preferences across different modes of autonomy (Omakase, Monitor, and Boss). The results revealed a preference for the active Boss mode, where participants took charge of crowd interactions, rather than relying entirely on the robot. Participants also appreciated the supportive role of the robot, using Monitor mode to understand the environment, negotiate movement, and facilitate user-robot interactions. These findings highlight the importance of shared control and user involvement for blind people. Studies on shared control for blind navigation are still neither active nor sufficient, requiring further research in these intersectional domains. Such studies will significantly expand the applicability of navigation robots in the future.

Acknowledgments

References

[1] Akira Akabayashi and RAYMOND Voltz. 2001. Advanced Directives in Different Cultures. *Topics in palliative care* 5 (2001), 107–122.
[2] Mauro Avila Soto, Markus Funk, Matthias Hoppe, Robin Boldt, Katrin Wolf, and Niels Henze. 2017. DroneNavigator: Using Leashed and Free-Floating Quadcopters to Navigate Visually Impaired Travelers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. ACM, New York, NY, USA, 300–304. https://doi.org/10.1145/3132525.3132556
[3] Minja Axelsson, Raquel Oliveira, Mattia Racca, and Ville Kyrki. 2021. Social robot co-design canvases: A participatory design framework. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 1 (2021), 1–39.
[4] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3–10.
[5] P Bachiller, D Rodriguez-Criado, RR Jorvekar, P Bustos, DR Faria, and LJ Manso. 2022. A graph neural network to model disruption in human-aware robot navigation. *Multimedia Tools and Applications* 81, 3 (2022), 3277–3295.
[6] Nikola Banovic, Rachel L. Franz, Khai N. Truong, Jennifer Mankoff, and Anind K. Dey. 2013. Uncovering information needs for independent spatial learning for users who are visually impaired. In *Proceedings of the 15th International ACM*

[7] SIGACCESS Conference on Computers and Accessibility (ASSETS '13). ACM, New York, NY, USA, Article 24, 8 pages. https://doi.org/10.1145/2513383.2513445
[7] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. ACM, New York, NY, USA, 161–173. https://doi.org/10.1145/3234695.3236348
[8] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. 2010. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 333–342.
[9] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934 [cs.CV]
[10] Stacy M Branham, Ali Abdolrahmani, William Easley, Morgan Scheuerman, Erick Ronquillo, and Amy Hurst. 2017. " Is Someone There? Do They Have a Gun" How Visual Information about Others Can Improve Personal Safety Management for Blind Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 260–269.
[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[12] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.
[13] Robin N Brewer and Vaishnav Kameswaran. 2018. Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. 185–197.
[14] Lucie Brunet, Françoise Darses, and Malika Auvray. 2018. Strategies and needs of blind pedestrians during urban navigation. *Le travail humain* 81, 2 (2018), 141–171.
[15] Dražen Brščić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from Children's Abuse of Social Robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)* (Portland, Oregon, USA). Association for Computing Machinery, New York, NY, USA, 59–66.
[16] Shaojun Cai, Ashwin Ram, Zhengtai Gou, Mohd Alqama Wasim Shaikh, Yu-An Chen, Yingjia Wan, Kotaro Hara, Shengdong Zhao, and David Hsu. 2024. Navigating Real-World Challenges: A Quadruped Robot Guiding System for Visually Impaired People in Diverse Environments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[17] Tara Capel and Margot Brereton. 2023. What is human-centered about human-centered AI? A map of the research landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. 1–23.
[18] Zaira Cattaneo and Tomaso Vecchi. 2011. *Blind vision: the neuroscience of visual impairment*. MIT press.
[19] Yanbo Chen, Zhengzhe Xu, Zhuozhu Jian, Gengpan Tang, Liyunong Yang, Anxing Xiao, Xueqian Wang, and Bin Liang. 2023. Quadruped Guidance Robot for the Visually Impaired: A Comfort-Based Approach. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. 12078–12084. https://doi.org/10.1109/ICRA48891.2023.10160854
[20] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. 2017. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1343–1350.
[21] Elizabeth R Chrastil and William H Warren. 2012. Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review* 19, 1 (2012), 23. https://doi.org/10.3758/s13423-011-0182-x
[22] Elizabeth R Chrastil and William H Warren. 2014. Active and Passive Spatial Learning in Human Navigation: Acquisition of Graph Knowledge. *Journal of experimental psychology. Learning, memory, and cognition* 41 (11 2014). https://doi.org/10.1037/xlm0000082
[23] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
[24] CMU-cabot. 2024. cabot. Retrieved in November 25, 2024 from https://github.com/CMU-cabot/cabot.
[25] Gonzalo Ferrer, Anaís Garrell, and Alberto Sanfeliu. 2013. Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1688–1694. https://doi.org/10.1109/IROS.2013.6696576
[26] Michael Fine and Caroline Glendinning. 2005. Dependence, independence or inter-dependence? Revisiting the concepts of 'care'and 'dependency'. *Ageing & society* 25, 4 (2005), 601–621.
[27] Anthony Francis, Claudia Pérez-d'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, et al. 2023. Principles and guidelines for evaluating social robot navigation algorithms. *arXiv preprint arXiv:2306.16740* (2023).

[28] Nora Fronemann, Kathrin Pollmann, and Wulf Loh. 2022. Should my robot know what's best for me? Human–robot interaction between user experience and ethical design. *AI & SOCIETY* 37, 2 (2022), 517–533.

[29] Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. 2023. What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–21.

[30] Nicholas A Giudice. 2018. Navigating without Vision: Principles of Blind Spatial Cognition. *Handbook of Behavioral and Cognitive Geography: Edward Elgar Publishing* (2018), 260–288. https://doi.org/10.4337/9781784717544.00024

[31] Nicholas A Giudice and Gordon E Legge. 2008. Blind navigation and the role of technology. *The engineering handbook of smart technology for aging, disability, and independence* (2008), 479–500.

[32] Jonathan Grudin. 2018. From tool to partner: The evolution of human-computer interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. 1–3.

[33] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M Kitani, and Chieko Asakawa. 2019. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *Proceedings of the 21st international ACM SIGACCESS Conference on Computers and Accessibility*. 68–82.

[34] Edward T Hall. 1963. A system for the notation of proxemic behavior. *American anthropologist* 65, 5 (1963), 1003–1026.

[35] Howard Ziyu Han, Franklin Mingzhe Li, Alesandra Baca Vazquez, Daragh Byrne, Nikolas Martelaro, and Sarah E Fox. 2024. Co-design Accessible Public Robots: Insights from People with Mobility Disability, Robotic Practitioners and Their Collaborations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. 1–20.

[36] Marion Hersh. 2015. Overcoming barriers and increasing independence–service robots for elderly and disabled people. *International Journal of Advanced Robotic Systems* 12, 8 (2015), 114.

[37] Jonggi Hong and Hernisa Kacorri. 2024. Understanding How Blind Users Handle Object Recognition Errors: Strategies and Challenges. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, Article 63, 15 pages. https://doi.org/10.1145/3663548.3675635

[38] Hochul Hwang, Hee-Tae Jung, Nicholas A Giudice, Joydeep Biswas, Sunghoon Ivan Lee, and Donghyun Kim. 2024. Towards Robotic Companions: Understanding Handler-Guide Dog Interactions for Informed Guide Dog Robot Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Article 596, 20 pages.

[39] Hochul Hwang, Tim Xia, Ibrahima Keita, Ken Suzuki, Joydeep Biswas, Sunghoon I. Lee, and Donghyun Kim. 2023. System Configuration and Navigation of a Guide Dog Robot: Toward Animal Guide Dog-Level Guiding Work. In *2023 IEEE International Conference on Robotics and Automation (ICRA '23)*. 9778–9784. https://doi.org/10.1109/ICRA48891.2023.10160573

[40] Gaurav Jain, Yuanyang Teng, Dong Heon Cho, Yunhao Xing, Maryam Aziz, and Brian A Smith. 2023. "I Want to Figure Things Out": Supporting Exploration in Navigation for People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1, Article 63 (2023), 28 pages. https://doi.org/10.1145/3579496

[41] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. 2019. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment* 37, 2 (2019), 140–153.

[42] Rie Kamikubo, Hernisa Kacorri, and Chieko Asakawa. 2024. "We are at the mercy of others' opinion": Supporting Blind People in Recreational Window Shopping with AI-infused Technology. *arXiv preprint arXiv:2405.06611* (2024).

[43] Rie Kamikubo, Naoya Kato, Keita Higuchi, Ryo Yonetani, and Yoichi Sato. 2020. Support strategies for remote guides in assisting people with visual impairments for effective indoor navigation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[44] Shaun K Kane, Chandrika Jayant, Jacob O Wobbrock, and Richard E Ladner. 2009. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. 115–122.

[45] Yuka Kaniwa, Masaki Kuribayashi, Seita Kayukawa, Daisuke Sato, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2024. ChitChatGuide: Enabling Exploration in a Shopping Mall for People with Visual Impairments through Conversational Interaction Using Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* MHCI (2024). https://doi.org/10.1145/3676492

[46] Seita Kayukawa, Keita Higuchi, João Guerreiro, Shigeo Morishima, Yoichi Sato, Kris Kitani, and Chieko Asakawa. 2019. BBeep: A sonic collision avoidance system for blind travellers and nearby pedestrians. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. 1–12.

[47] Seita Kayukawa, Daisuke Sato, Masayuki Murata, Tatsuya Ishihara, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2023. Enhancing Blind Visitor's Autonomy in a Science Museum Using an Autonomous Navigation Robot. In

Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23) (Hamburg, Germany). ACM, New York, NY, USA, Article 541, 14 pages. https://doi.org/10.1145/3544548.3581220

[48] Kyoko Konishi and Véronique Bohbot. 2013. Spatial navigation strategies correlate with gray matter in the hippocampus of healthy older adults tested in a virtual maze. *Frontiers in aging neuroscience* 5 (2013), 1. https://doi.org/10.3389/fnagi.2013.00001

[49] Masaki Kuribayashi, Tatsuya Ishihara, Daisuke Sato, Jayakorn Vongkulbhisal, Karnik Ram, Seita Kayukawa, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2023. PathFinder: Designing a Map-Less Navigation System for Blind People in Unfamiliar Buildings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)* (Hamburg, Germany). ACM, New York, NY, USA, Article 41, 16 pages. https://doi.org/10.1145/3544548.3580687

[50] Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2021. Linechaser: a smartphone-based navigation system for blind people to stand in lines. In *Proceedings of the 2021 CHI conference on human factors in computing systems (CHI '21)*. 1–13.

[51] Sooyeon Lee, Madison Reddie, and John M Carroll. 2021. Designing for Independence for people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–19.

[52] Christina Lichtenthäler, Tamara Lorenzy, and Alexandra Kirsch. 2012. Influence of Legibility on Perceived Safety in a Virtual Human-Robot Path Crossing Task. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 676–681.

[53] Shuijing Liu, Peixin Chang, Weihang Liang, Neeloy Chakraborty, and Katherine Driggs-Campbell. 2021. Decentralized Structural-RNN for Robot Crowd Navigation with Deep Reinforcement Learning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 3517–3524.

[54] Steve Macenski, Matthew Booker, and Josh Wallace. 2024. Open-Source, Cost-Aware Kinematically Feasible Planning for Mobile and Surface Robotics. *Arxiv* (2024).

[55] Steve Macenski, Tom Moore, David V Lu, Alexey Merzlyakov, and Michael Ferguson. 2023. From the desks of ROS maintainers: A survey of modern & capable mobile robotics algorithms in the robot operating system 2. *Robotics and Autonomous Systems* 168 (2023), 104593.

[56] Nihongo Master. 2024. Japanese Dictionary - Omakase. https://www.nihongomaster.com/japanese/dictionary/word/100961/omakase-%E3%81%8A%E4%BB%BB%E3%81%9B-%E5%BE%A1%E4%BB%BB%E3%81%9B-%E3%81%8A%E3%81%BE%E3%81%8B%E3%81%9B?utm_term=word&utm_content=100961&utm_source=copylink&utm_medium=sharer&utm_campaign=sharer

[57] Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dapeng Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. 2023. Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–39.

[58] A Allan Melvin, B Prabu, R Nagarajan, and Bukhari Illias. 2009. ROVI: A Robot for Visually Impaired for Collision-free Navigation. In *Proceedings of the International Conference on Man-Machine Systems (ICoMMS 2009)*. Universiti Malaysia Perlis, Kangar and Arau, Perlis, Malaysia, 3B5–1.

[59] Stephanie M Merritt, Alicia Ako-Brew, William J Bryant, Amy Staley, Michael McKenna, Austin Leone, and Lei Shirase. 2019. Automation-induced complacency potential: Development and validation of a new scale. *Frontiers in psychology* 10 (2019), 225.

[60] James W Moore. 2016. What is the sense of agency and why does it matter? *Frontiers in psychology* 7 (2016), 1272.

[61] Yasushi Nakauchi and Reid Simmons. 2000. A social robot that stands in line. In *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)*, Vol. 1. 357–364 vol.1.

[62] Nippon.com. 2020. Japan Struggling to Increase the Number of Active Guide Dog. https://www.nippon.com/en/japan-data/h00644/japan-struggling-to-increase-the-number-of-active-guide-dogs.html

[63] OpenAI. 2024. Hello GPT-4o | OpenAI. Retrieved in September 8, 2024 from https://openai.com/index/hello-gpt-4o/.

[64] Vinitha Ranganeni, Mike Sinclair, Eyal Ofek, Amos Miller, Jonathan Campbell, Andrey Kolobov, and Edward Cutrell. 2023. Exploring levels of control for a navigation assistant for blind travelers. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. 4–12.

[65] Patricia J Rock. 1988. Independence: What it means to six disabled people living in the community. *Disability, Handicap & Society* 3, 1 (1988), 27–35.

[66] Alberto Rodríguez, J Javier Yebes, Pablo F Alcantarilla, Luis M Bergasa, Javier Almazán, and Andrés Cela. 2012. Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. *Sensors* 12, 12 (2012), 17476–17496.

[67] Sushil Kumar Sahoo and Bibhuti Bhusan Choudhury. 2024. Autonomous navigation and obstacle avoidance in smart robotic wheelchairs. *Journal of Decision Analytics and Intelligent Computing* 4, 1 (2024), 47–66.

[68] Supraja Sankaran, Chao Zhang, Marisela Gutierrez Lopez, and Kaisa Väänänen. 2020. Respecting human autonomy through human-centered AI. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–3.

[69] Kai Shen, Zeqian Ju, Xu Tan, Eric Liu, Yichong Leng, Lei He, Tao Qin, sheng zhao, and Jiang Bian. 2024. NaturalSpeech 2: Latent Diffusion Models are Natural and Zero-Shot Speech and Singing Synthesizers. In *The Twelfth International Conference on Learning Representations*.

[70] Kristen Shinohara and Jacob O Wobbrock. 2016. Self-conscious or self-confident? A diary study conceptualizing the social accessibility of assistive technology. *ACM Transactions on Accessible Computing (TACCESS)* 8, 2 (2016), 1–31.

[71] Ben Shneiderman. 2022. *Human-centered AI.* Oxford University Press.

[72] Brian T Slingsby. 2004. Decision-making models in Japanese psychiatry: transitions from passive to active patterns. *Social science & medicine* 59, 1 (2004), 83–91.

[73] Susumu Tachi and Kiyoshi Komoriya. 1985. Guide dog robot. In *Proceedings of the Second International Symposium on Robotics Research (1984).* MIT Press, Cambridge, MA, USA, 333–340.

[74] Phani Teja Singamaneni, Anthony Favier, and Rachid Alami. 2021. Human-Aware Navigation Planner for Diverse Human-Robot Interaction Contexts. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* 5817–5824.

[75] Kazuteru Tobita, Katsuyuki Sagayama, and Hironori Ogawa. 2017. Examination of a guidance robot for visually impaired people. *Journal of Robotics and Mechatronics* 29, 4 (2017), 720–727.

[76] Peter Trautman and Andreas Krause. 2010. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems.* 797–803.

[77] Michael Tucker and Sheila Hocken. 1984. The Eyes That Lead: The story of guide dogs for the blind. *(No Title)* (1984).

[78] Iwan Ulrich and Johann Borenstein. 2001. The GuideCane-applying mobile robot technologies to assist the visually impaired. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* 31, 2 (2001), 131–136. https://doi.org/10.1109/3468.911370

[79] Andreas Wachaja, Pratik Agarwal, Mathias Zink, Miguel Reyes Adame, Knut Möller, and Wolfram Burgard. 2017. Navigating blind people with walking impairments using a smart walker. *Autonomous Robots* 41, 3 (01 Mar 2017), 555–573. https://doi.org/10.1007/s10514-016-9595-8

[80] Allan Wang, Christoforos Mavrogiannis, and Aaron Steinfeld. 2022. Group-based motion prediction for navigation in crowded environments. In *Conference on Robot Learning.* PMLR, 871–882.

[81] Allan Wang and Aaron Steinfeld. 2020. Group Split and Merge Prediction With 3D Convolutional Networks. *IEEE Robotics and Automation Letters* 5, 2 (2020), 1923–1930.

[82] Yuanlong Wei, Xiangxin Kou, and Min Cheol Lee. 2014. A new vision and navigation research for a guide-dog robot system in urban system. In *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM '14).* IEEE Press, Piscataway, NJ, USA, 1290–1295. https://doi.org/10.1109/AIM.2014.6878260

[83] William R Wiener, Richard L Welsh, and Bruce B Blasch. 2010. *Foundations of orientation and mobility.* Vol. 1. American Foundation for the Blind.

[84] Michele A Williams, Caroline Galbraith, Shaun K Kane, and Amy Hurst. 2014. " just let the cane hit it" how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers and Accessibility.* 217–224.

[85] Michele A Williams, Amy Hurst, and Shaun K Kane. 2013. " Pray before you step out" describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility.* 1–8.

[86] Nancy Worth. 2013. Visual impairment in the city: Young people's social strategies for independent mobility. *Urban Studies* 50, 3 (2013), 574–586.

[87] Xuesu Xiao, Tingnan Zhang, Krzysztof Marcin Choromanski, Tsang-Wei Edward Lee, Anthony Francis, Jake Varley, Stephen Tu, Sumeet Singh, Peng Xu, Fei Xia, Sven Mikael Persson, Dmitry Kalashnikov, Leila Takayama, Roy Frostig, Jie Tan, Carolina Parada, and Vikas Sindhwani. 2023. Learning Model Predictive Controllers with Real-Time Attention for Real-World Navigation. In *Conference on Robot Learning.* PMLR, 1708–1721.

[88] Bingqing Zhang, Giulia Barbareschi, Roxana Ramirez Herrera, Tom Carlson, and Catherine Holloway. 2022. Understanding interactions for smart wheelchair navigation in crowds. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–16.

[89] Yan Zhang, Ziang Li, Haole Guo, Luyao Wang, Qihe Chen, Wenjie Jiang, Mingming Fan, Guyue Zhou, and Jiangtao Gong. 2023. "I am the follower, also the boss": Exploring Different Levels of Autonomy and Machine Forms of Guiding Robots for the Visually Impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* 1–22.