# Teaching a Robot Where to Park: A Scalable Crowdsourcing Approach

De'Aira Bryant[1,*], Tiago Etiene[2], Ayanna Howard[1,3], William D. Smart[2,4], Dylan F. Glas[2]

***Abstract*— For social robots to successfully integrate into daily life in home environments, they will need reliable models of the way people perceive and use space in the home. This paper explores the problem of obtaining annotated training data at scale for subjective judgments about spatial locations. Focusing on the use case of identifying good and bad parking spots for a social robot operating in a home environment, two experiments are presented. The first study shows that the presentation of context-rich 3D images to human annotators yields notably different outcomes from those obtained when using 2D robot navigation maps. We attribute the source of these differences to a set of features visible only in the 3D views and introduce a technique for labeling these features on the 2D maps. The second study reveals that using labeled 2D maps produces annotation data very similar to that obtained using 3D images. Since a labeled 2D map can be generated at a fraction of the cost of a full set of 3D views, we recommend this method as a scalable approach to collecting subjective spatial data annotations in everyday environments.**

## I. INTRODUCTION

Social robots have become increasingly popular in venues that prompt short-term interactions with humans like hotels and grocery stores, but they have yet to to convince users of their adaptability and usefulness in venues that require long-term interactions like people's homes [1]. Research in innovation adaptation suggests that these robots will need to display intelligent and social behavior to capture the interest of mass markets and potential home users [2].

Taking on the challenge, Amazon has recently released Astro, a household robot for home monitoring (see Fig. 1) [3]. One of Astro's intelligent motion features allows it to find its way around a home and hang out near users at the ready while not in use. Astro needs to choose hangout parking locations that are out of the way to avoid obstructing walking paths or items in the home the users might need to access. These are long-term parking spots where Astro may be parked for several hours. To achieve this goal effectively, Astro must *understand and model the way people perceive and use space in the home*.

[1]De'Aira Bryant and Ayanna Howard are affiliated with the College of Computing at Georgia Institute of Technology.

[2]Tiago Etiene, William D. Smart, and Dylan F. Glas are affiliated with Amazon Lab126.

[3]Ayanna Howard is also affiliated with the College of Engineering at The Ohio State University.

[4]William D. Smart is also affiliated with the Collaborative Robotics and Intelligent Systems (CoRIS) Institute at Oregon State University.

[*]Experimental work performed during an Amazon Lab126 internship.

The emails of authors are: dbryant@gatech.edu, tiagoq@amazon.com, howard.1727@osu.edu, smartw@oregonstate.edu, dgglas@amazon.com

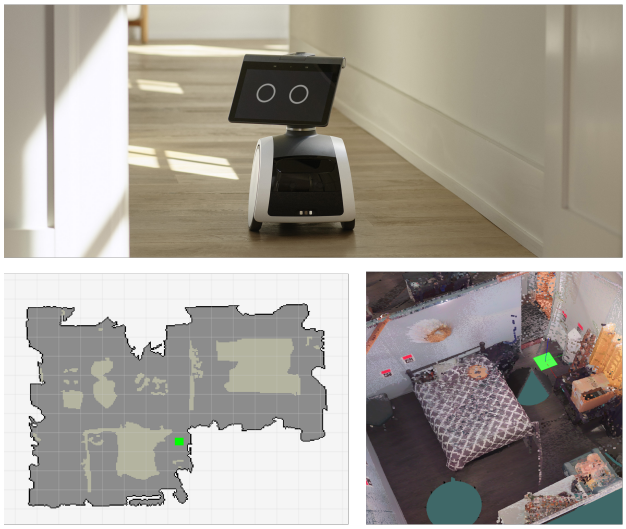

Fig. 1. Astro is Amazon's new household social robot. In this paper, we explore human perceptions of social appropriateness of Astro parking in different locations in the home using two visual representations. Here, we show examples of the dataset stimuli for the 2D and 3D data shown on the bottom left and right respectively.

Human spatial preferences often rely on contextual, cultural, and subjective judgments and therefore would be difficult to capture using simple modeling techniques. In this work, we propose to collect human judgments of spatial preference across multiple home environments. This data can be used as ground truth to train and validate more complex models of home spatial use. Such models can then be used to develop and validate behaviors for robots like Astro that rely on such knowledge. To support Astro's hangout behavior, we want to model where humans think it should park. We focus on the specific problem of the *social appropriateness of parking in different locations in the home*.

Asking home owners to directly annotate all good and bad parking spots in their homes is labor-intensive. Professional analysis teams could visit test homes to collect this data, but this method would be difficult to scale and generalize to the diversity of different home environments. To address this limitation, we propose crowdsourcing as a method to collect these human judgments quickly and efficiently. However, quality crowdsourced responses rely on having an accurate representation of the situated environment [4].

In this paper, we consider two possible representations of Astro's environment to generate data for crowdsourcing. The first representation is a two-dimensional (2D) top-down view of a home's layout generated from Astro's navigational map. The 2D data requires minimal effort to generate but

does not capture all the context necessary for reliable human judgments. The second representation is a three-dimensional (3D) RGB point-cloud image of the home generated from a separate lidar scan. The 3D data provides a richer interpretation of the home but requires substantial effort and expensive equipment to generate.

We will first describe a preliminary user study conducted to determine the potential for using these two representations to collect human judgments of the social appropriateness of Astro's parking locations. We sought to understand whether 2D maps were sufficient for collecting annotations of robot parking spots, or whether the richer information available in 3D images would be necessary. Our preliminary results informed the design of a second user study that included an intermediate representation that provides additional context to the 2D maps and reduces the effort and computational power required to generate the dataset stimuli for annotators. This approach is shown to facilitate a more efficient and scalable data collection process for collecting and validating robot location features derived from subjective spatial understanding of the home.

## II. Related Work

### A. Expectations for Home Robots

Social robots have been considered for use in various tasks in the home environment ranging from care [5], [6] to assistance [7], [8] to companionship [9], [10], and even entertainment [11]. Graaf found that increasing the robot's sociability is one way to foster the acceptance of social robots in home environments. Recommendations to increase sociability include enhanced conversational abilities, user detection and recognition, emotional behavior and perception, and intelligent navigational capabilities [12]. While prior work has considered robot navigation and path planning for social robots in the home when the target parking location is known or provided by users [13], [14], very little work has considered how the target location can be chosen autonomously. In this work, we propose to teach social robots which parking locations are appropriate by crowdsourcing human judgments across multiple home floorplans.

### B. Social Positioning & Spatial Understanding

The social positioning of mobile robots has largely been considered in the context of interactions with one or more humans in a given environment [15], [16]. Approaches to determine the appropriate social positioning include direct instruction from users [17], inference from non-verbal social cues [16], and prior knowledge related to the robot's objectives [18]. Social positioning has also been explored in a diverse set of environments, including shopping malls [19], nursing homes [18], and emergency rooms [20] to name a few. Many of these works collectively agree that social robots should remain out of the way for humans that are focused on completing tasks in the same collocated space. Thus raises the question, how does a robot learn to stay "out of the way"? This complex spatial characteristic requires an underlying understanding of how humans use space in the home. However, little work has attempted to model spatial understanding of the home to help determine where a robot should position itself. Our approach will demonstrate that appropriate social robot positions can be measured efficiently even for complex environments like people's homes.

## III. 2D vs. 3D Preliminary Study

An accurate spatial representation of a potential parking position in a given home environment is needed to crowdsource preference data from human annotators. We began our exploration by considering a 2D and a 3D representation that were readily available. Once deployed, Astro uses advanced navigational technology to find its way around the home environment and generates 2D navigational maps. 3D point-cloud representations of the home can also be collected by setting up standalone lidar sensors. The coordinate systems of these 2D and 3D home representations can be aligned such that a potential robot position can be examined in both worlds. We used a selection of these maps and scans to produce a set of stimuli for a preliminary crowdsourcing experiment where we sought to gather human judgements of social appropriateness of robot parking locations.

### A. 2D Home Representations (2D-U)

The 2D maps illustrate the navigable space in a given home environment and represent the world as Astro sees it. An example map is shown on the bottom left in Fig. 1. Astro builds the navigational 2D map using SLAM for every environment it operates in [21]. As such, these maps are the most easily accessible and prevalent source of home spatial configurations we have. They most closely resemble the architectural floor plans that people would use to examine the layout of homes or apartments. However, our 2D navigational maps only display walls and "clutter". Household features like doors, sofas and room names are not labeled explicitly.

For our 2D image dataset stimuli, we uniformly sampled potential robot parking spots across the navigable area in each map generated from exploration of different home environments. A potential parking spot for the robot is shown as a bright green square (0.4m per side, about the size of the physical Astro robot) plotted on the map (see Fig. 1).

### B. Immersive 3D Home Representations

The 3D RGB lidar scans illustrate the home environment as humans would see it. An example view is shown on the bottom right in Fig. 1. The lidar scans were collected using standalone Leica BLK360 scanners. From the lidar scans, we produced detailed full-color point clouds, showing rich contextual information about the home to help inform annotator ratings. Collection of these 3D scans required additional effort, expensive equipment, and specialized software.

To generate the 3D image dataset stimuli, we uniformly sampled potential robot parking spots across the navigable floor space. The potential parking spot is indicated on each captured image with a bright green square plotted on the floor (about the relative size of the physical Astro robot, about 0.4m per side). We needed to check and potentially

adjust the camera angle *per potential parking location* to best capture the pose and its relevant context. For example, Fig. 2 shows two potential parking locations that are near each other but required the manual adjustment of the camera angle to visibly see each location. This camera-adjustment process required special software to manipulate the 3D point-clouds.

### C. Robot Pose Annotation Collection

The 2D and 3D image dataset stimuli of the home environments were used to collect human judgements of the appropriateness of potential robot parking locations. We designed an annotation interface to allow raters to easily access the task and allow us to quickly process the results.

*1) Annotation Interface:* Amazon SageMaker Ground Truth was used to create the user interface. The interface provided instructions for annotators in the left panel, displayed the experimental stimuli in the middle panel, and showed the robot parking location rating scale in the right panel (Fig. 3). Annotators had the ability to use shortcuts to provide ratings, zoom and pan the image stimuli, and review the instructions or rating scale guidance at any time.

*2) Annotator Instructions:* Annotators were first told that the purpose of the annotation task was to help model the quality of locations for a robot to park in various home environments. They were then instructed to watch a short two-minute video introducing Amazon Astro [3]. Instructions were then provided that described the objective of the robot, illustrated examples of decisively good and bad robot parking locations, and provided guidance for interpreting the image stimuli they would be rating (i.e., the green parking spot, the relative size of the grid, etc.) and the scale to use for evaluating the potential parking location. Finally, they were given tips on using the interface efficiently and instructed to spend no more than one hour at a time providing annotations.

*3) Robot Pose Rating Scale:* Annotators were instructed to rate each parking spot using the following scale:

1) **Very Poor** - Astro absolutely must not park here.
2) **Poor** - Astro should try to avoid this parking spot. (e.g. users could walk around Astro but it is inconvenient)
3) **Cannot Determine** - Not clearly bad, but also not clearly good.

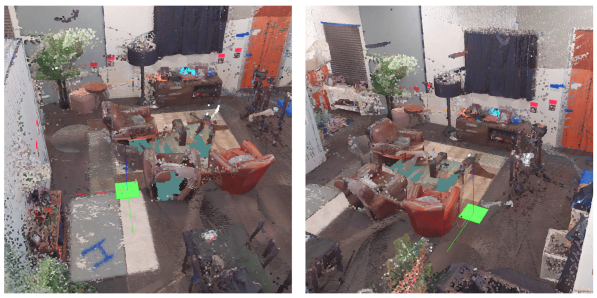

Fig. 2. Two nearby parking positions that cannot be seen from the same perspective, requiring manual adjustment of the camera angle to be seen. This step, requiring specialized software and manual effort, is conducted per potential parking position and may need to be repeated hundreds of times to generate the dataset stimuli needed for crowdsourcing.

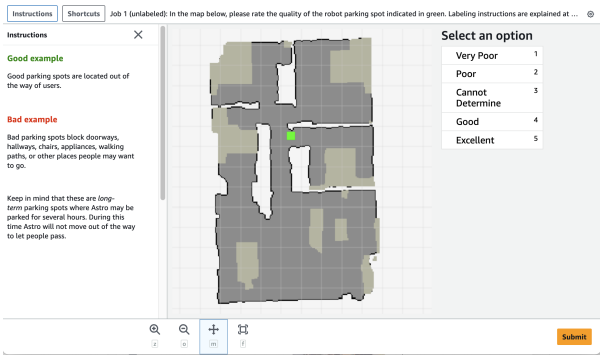


Fig. 3. The developed interface for annotators to provide subjective judgements of appropriateness for robot parking. The interface was used to annotate the image stimuli during each condition for each experiment.

4) **Good** - Astro can park here but it is not perfect.
5) **Excellent** - It is ideal for Astro to park here.

### D. Participants

In accordance with the Amazon company policy, data annotation was performed by a paid internal annotation team. Rather than using a public crowdsourcing platform such as Mechanical Turk, a contracted annotation team was used to ensure compliance with relevant data privacy protocols. Four annotators were selected to provide social appropriateness ratings for each potential parking location across all floor plans. Participants had no direct experience with the Astro home robot.

### E. Experimental Stimuli

In our preliminary user study, we utilized a total of 4 home floor plans obtained from a dataset of non-customer home floor plans which were internally collected for research purposes. We sampled for potential parking locations across the navigable area of each floor plan at a spacing of 0.5 meters, resulting in 562 locations. The datasets of image stimuli were then generated where each of the 562 total potential parking locations were shown individually on their respective floor plans. In other words, 2D map views of potential parking locations were directly compared to 3D image views of those same parking locations. All annotators completed the 2D annotation task before the 3D task on separate days to avoid carryover effects.

### F. Results

We first examined the data to determine the level of variance between annotator ratings. The average standard deviation across all the sample poses was 0.68 for the 2D image ratings and 0.60 for the 3D image ratings. The mean absolute deviation between ratings of corresponding poses for the two conditions was 0.93. Raters agreed perfectly on 102/562 (18.1%) of poses. Annotators were in slightly better agreement in the 3D condition than the 2D.

Next, we explored what features of the home contributed to the most variance between annotator ratings. We define an annotator social appropriateness rating as $r^d_{i,j} \in$

$\{0, 0.25, 0.5, 0.75, 1\}$, where $d \in \{2,3\}$ is the dimension viewed to obtain the rating, $i \in [1,2,...,N]$ indicates the sample parking position index ($N = 562$) of the image stimuli, and $j \in [1,2,...,M]$ indicates the human annotator index ($M = 4$). We then define an aggregate social appropriateness rating for each sample $i$ per dimension $d$ as

$$S_{d,i} = \frac{1}{M} \sum_{j=1}^{M} r_{i,j}^{d} \quad (1)$$

We then inspected the aggregate differences between the annotator's ratings of the 2D and 3D presented parking positions. We calculate the absolute distance between each aggregate rating $A_i = |S_{3,i} - S_{2,i}|$ for each sample parking position $i$. A difference map for one of the four floor plans is illustrated in Fig. 4. The values shown are the absolute distance scores $A_i$ and the color of the parking positions indicate which condition had lower social appropriateness ratings and were thus more strict. Red locations indicate the 3D ratings were more strict and blue locations indicate the 3D ratings were more tolerant toward Astro parking there. Most frequently, ratings were more strict in the 3D condition. In a few cases, the trends went in the other direction. In all cases, the cause of the discrepancy was easily determined by inspection. For example, some poses blocked doors which were visible in the 3D view but not on the 2D map. Others blocked access to appliances or high-traffic areas in the kitchen, something also not visible on the 2D map.

We performed a qualitative examination on clusters of potential parking locations, focusing on areas where the average scores differed by at least two points between conditions (i.e., where annotators contradicted their rating in the alternate condition). Most frequently, these clusters coincided with one of the following features: kitchens, doors, beds, tables, and sofas. All of these features are explicitly visible in the 3D images, but not in the 2D maps. Of 562 total potential parking positions, 91 (16.2%) had a rating difference of at least 2.0 between the 2D and 3D image rating. Of these, only two received more lenient ratings in the 3D condition. The frequency of features that contributed to the discrepancy in annotator ratings is shown in Fig. 5.

### G. Discussion

Our preliminary results indicate that annotations differed between the 2D maps and 3D images most where kitchens, doors, beds, and other furniture could not be effectively interpreted in the 2D maps. The 3D presented parking positions resulted in more consistent annotator ratings, but the process to generate the 3D parking positions is both tedious and time-intensive where each potential parking position must be individually captured per map, presenting a huge scalability issue. A more scalable approach that provides the necessary features and contexts of the home environment is needed to measure social appropriateness of parking positions effectively.

The conclusion from this preliminary user study was that much of the discrepancy between the 2D and 3D conditions

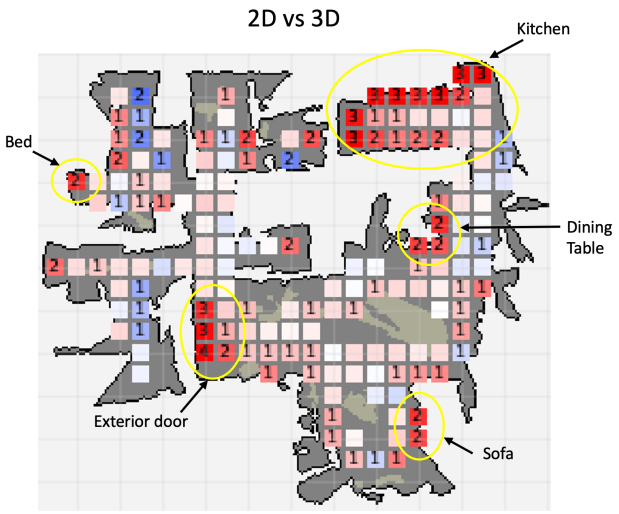


Fig. 4. A visualization showing the difference between aggregate annotator ratings for individual positions in the 2D and 3D conditions. Intensity of color indicates magnitude of differences between conditions. Red areas were rated more strictly in 3D and blue areas more leniently. Numbers indicate absolute difference for positions with differences of at least 0.5 points.

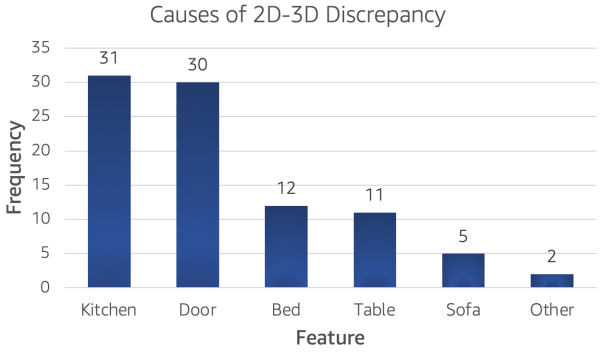


Fig. 5. Observed frequency of features that contributed to discrepancies between annotator's social appropriateness ratings of 2D and 3D-presented potential parking locations. Kitchens, doors, beds, tables, and sofas were the most common sources of discrepancy.

was explained by a small number of contextual features which are visible in the 3D images. This finding suggested that if these features were labeled in the 2D map, we might be able to achieve similar performance to the 3D condition. The following section describes a second user-study we conducted to explore this by comparing annotator ratings of 2D, labeled 2D, and 3D presented parking positions.

## IV. Labeled 2D Maps User Study

Results from the preliminary user study suggested that the original 2D maps did not model enough of the home environment's context for annotators to provide useful and consistent ratings. The high-fidelity 3D image stimuli resulted in improved annotator performance, but the process to generate the samples was not efficient. Based on these findings, we hypothesized that adding labels of important features to the 2D map might achieve the annotation accuracy


Fig. 6. An example 2D-L map. Contextual information is explicitly identified to help annotators provide an informed rating of social appropriateness for a given potential robot parking position. Light brown areas represent clutter or furniture, red arcs or line segments indicate doors, colored boxes indicate kitchen, other rooms, or beds, and the remainder of features are labeled with text. 1-meter grid lines provide scale reference.

of the 3D condition, while maintaining the scalability and lower data generation cost of the 2D condition. We conducted a user study to test this hypothesis comparing unlabeled 2D maps (2D-U), labeled 2D maps (2D-L) and 3D images (3D) of multiple home environments. The 2D-U maps were described in Section III-A, the 3D images in Section III-B, and the 2D-L maps are described next.

### A. Labeled 2D Maps (2D-L)

We generated 2D-L from the maps introduced in Section III-A. An additional data-preparation step was applied where we explicitly labeled contextual features on the maps used to generate the image stimuli. All of the features discovered in the previous study were added: kitchens, doors, beds, sofas, and tables. Additional features were added to some maps, including toilets, staircases, chairs, desks, and fireplaces. Many of these features were not discernible in the initial 2D maps (2D-U) and our goal was to provide an intuitive description of major features in the home. We also determined that it was necessary to specify the swing arc of each door, as this provided important information regarding potential parking spaces in the vicinity of the door.

An example 2D-L map is shown in Fig. 6. These maps were created manually, by overlaying text and shapes on the robot's navigation maps. Information for the map annotation came from manual inspection of the 3D point clouds, but it would also be possible to gather this information from in-person inspection, architectural floor plans, or photos of the environment, depending on the method of collecting the map data. This step is only needed once per map, thus making the process more scalable and efficient than the 3D approach.

### B. Experimental Stimuli

For experimentation, we utilized a new set of 6 non-customer floor plans and uniformly sampled potential parking positions at a spacing of 0.5 meters across navigable space. In this data set, some floor plans were much larger than others, so to prevent the larger floor plans from dominating the dataset, the number of poses was sub-sampled to a maximum of 200 per floor plan. We then selected 1,000 total potential parking locations for inclusion in the experiment. We then generated the 3 datasets of image stimuli for 2D-U, 2D-L, and 3D representational approaches. We utilized the same annotation collection technique from our preliminary user study as described in Section III-C.

In the previous study it was clear that the 3D condition provided more information than the 2D-U condition, making the ordering of conditions straightforward. In this study, the 2D-U condition was again presented first, but it was not clear whether the 2D-L or 3D condition would provide a greater amount of semantic information. While the 3D view showed richer information, it was also noisy and sometimes difficult to interpret. The 2D-L map had clear labels but was missing fine details. In consideration of this ambiguity, the 2D-L and 3D conditions were counterbalanced. For half of the floorplans the 2D-L condition was presented first, and for the remainder, the 3D condition was presented first.

### C. Participants

We employed the same 4 professional annotators as in the previous experiment. Each annotator provided social appropriateness ratings for each of the 1,000 potential parking locations for each of the 2D-U, 2D-L, and 3D annotation tasks, resulting in a total of 12,000 annotations.

### D. Results

*1) Inter-Rater Reliability:* We again examined the data to determine the level of variance between annotator ratings. The average standard deviation across all the sample poses was 0.90, 0.89, and 0.98 for the 2D-U, 2D-L and 3D ratings respectively. Next, we calculated the average inter-rater reliability between the 4 annotators using Cohen's Kappa with quadratic weighting ($\kappa$). This increases the "badness" of a disagreement quadratically with the size of the disagreement, and is useful for ordinal scales such as the one we are using. The $\kappa$ score for the 2D-U, 2D-L, and 3D ratings was 0.42, 0.45 and 0.34 respectively, signifying fair agreement for the 2D-U and 2D-L conditions, and slight agreement for the 3D condition. Average agreement was strongest on the 2D-L data, and weakest on the 3D data. The labeling of the 2D-U maps gives a consistent set of semantics to the map, which might result in less disagreement. The average agreement for the 2D-U data was Fair, suggesting that the annotators are consistent in their treatment of open space and features interpreted as walls. Further inspection of individual annotator ratings show that annotator pairs did not seem to consistently agree or disagree across the conditions. We attribute the observed variance to subjective differences in perception and utilization of space in the home. These results support that multiple judgements are needed to capture the range of spatial preference amongst potential human users.

*2) Inter-Condition Variation:* Next, we inspected the aggregate differences between the annotator's ratings across the three conditions. We again calculated the absolute difference between conditions for the aggregate ratings for each sample parking position (see Section III-F). Fig. 7 visualizes the

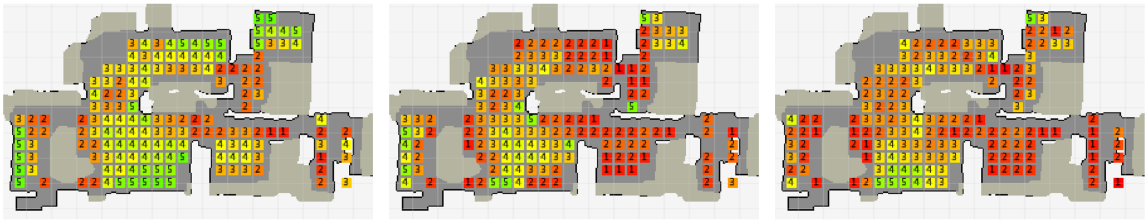

Fig. 7. A visualization of annotators' social appropriateness ratings for the 2D-U (Left map), 2D-L (Center map) and 3D (Right map) presented parking positions. Here, we can see all the sampled parking positions for the given floor plan to determine the general areas and spots that are deemed acceptable for Astro. Green potential parking positions had higher annotator ratings than red parking positions.

aggregate ratings for a sample floor plan under each of the three conditions. Fig. 8 shows difference maps between each pair of conditions. By inspection it is clear that both the 3D and 2D-L conditions resulted in stricter annotations overall, compared with the 2D-U baseline, particularly in areas like the kitchen and near doors. This finding is consistent with trends observed in the previous study. Fig. 8 (right) shows the mean differences between ratings, indicating that the 2D-L condition produced scores much closer to the 3D condition than 2D-U did.

In Table I shows relevant statistics to provide insight into the annotator behavior between conditions. We report (1) the RMS for the two conditions; (2) the mean difference between the two conditions; (3) the Spearman's r correlation statistic between the ratings ($\rho$); (4) the percentage of locations that did not change rating between the evaluations (*const*); and (6) the percentage of locations that did not change rating by more than one point, in either direction ($const_1$). The average RMS difference in scores is smallest between the 3D and the 2D-L maps (1.33), and largest between the 2D-U and 3D maps. The difference between the 2D-U and the 2D-L maps is similarly large (1.62), suggesting that both the 2D-L and the 3D maps are conveying information that is being interpreted more consistently than the 2D-U map. Despite the non-normality of the data, a Student's t-test is appropriate given the large number of annotated hangout poses. We found no significant difference between the means of the ratings when considering all of the annotations (at 0.05 level). However, the actual mean values are quite close to each other and could easily have been dominated by the density of poses in open spaces. As such, we interpret the higher correlation between scores in the 2D-L and 3D conditions, coupled with a low (0.11) mean change in scores, to suggest that the data in the 2D-L dataset is being interpreted similarly to that in the 3D. The number of points that remained constant across conditions is highest for the 2D-L and 3D conditions, further suggesting that they are being interpreted most consistently. Overall observations across all 6 maps indicate that annotator ratings were most similar for the 2D-L and 3D presented parking positions. The ratings are not normally distributed (as verified by a Shapiro-Wilks test, with all $p < 0.001$), so testing for differences between the conditions with an ANOVA is not appropriate. However, a Kruskal-Wallace test revealed that there was a significant difference in medians across both conditions and annotators for most of the floorplans (all except one for condition, and two for annotator), at the 0.05 significance level. This suggests that the condition (2D-U, 2D-L, 3D) causes a different annotation and also that what constitutes a good place for the robot to wait is a very subjective thing.

| Condition | RMS | mean | $\rho$ | const | const_1 |
|---|---|---|---|---|---|
| **2D-U - 3D** | 1.72 | 0.68 | 0.27 | 29% | 63% |
| **2D-U - 2D-L** | 1.62 | 0.57 | 0.32 | 32% | 67% |
| **2D-L - 3D** | 1.33 | 0.11 | 0.42 | 51% | 75% |

TABLE I

HERE, WE COMPARE THE INTER-CONDITION VARIATION BETWEEN ANNOTATION RATINGS OF SOCIAL APPROPRIATENESS.

*3) Rating Discrepancy Analysis:* Next, we again examined the potential parking positions with a rating difference of at least 2.0 between the 3D and 2D-L conditions. In total, there were 31 such poses across the 6 floor plans, constituting only 3.1% of the data set. Table II shows our best assessment of the relative frequency of detected issues. The most prevalent issues were doors behind curtains and the perception of open space. Glass doors behind curtains were not easily recognized as doors in the 3D view, and in a few cases, doors were assumed to be present which were not really there. This suggests a benefit to the 2D-L map, where labels of door positions can be carefully checked and explicitly indicated. In some cases, it seems that the height of furniture influenced the perception of openness of space around it. In other cases, this difference in perception may have been due to inaccuracy of the robot's navigational map. This would suggest a benefit to the 3D visualization in terms of conveying a more accurate sense of space and scale. These examples suggest there is an overall advantage to the 2D-L approach over 3D, although the results are mixed. These observations may be useful for improving the method further, but it should be emphasized that the issues examined here are infrequent and only reflect 3.1% of the data.

## V. DISCUSSION

Insights from our preliminary experiment led us to hypothesize that potential parking positions shown on 2D-L maps

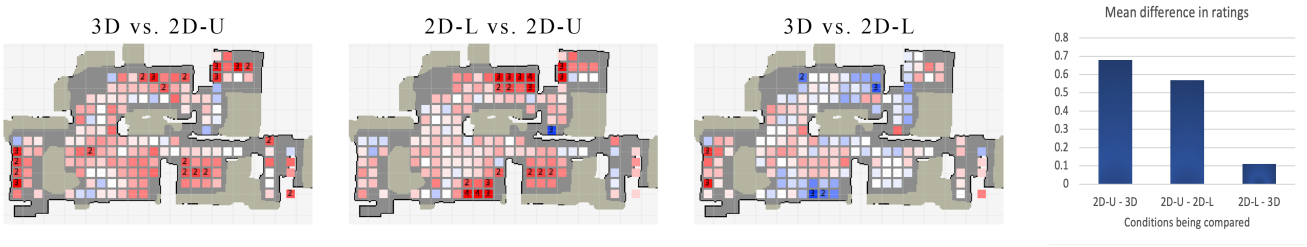


Fig. 8. Left: Difference maps between average ratings in the experimental conditions. Red indicates that annotations in the former condition were stricter than in the latter, and blue indicates that they were more lenient. Darker squares with numbers in them represent differences $\geq 2.0$ rating points. Right: Mean annotation score differences between conditions. Mean difference between 2D-L and 3D scores is much lower than either condition vs. 2D-U.

| Issue | Count | Advantage |
|---|---|---|
| Doors behind curtains | 7 | 2D-L |
| Seems more open in 2D | 7 | 3D |
| Seems more open in 3D | 2 | 3D |
| Map noise / boundary unclear in 2D | 5 | 2D-L |
| Pixel noise / boundary unclear in 3D | 4 | 2D-L |
| 3D perspective misleading | 3 | 2D-L |
| Chair/Sofa orientation unclear in 2D-U | 2 | 3D / Correctable |
| Label missing | 1 | 3D |
| **Total** | 31 | |

TABLE II

LIKELY CAUSES OF NOTABLE DISCREPANCY BETWEEN THE 3D AND 2D-L ANNOTATOR RATINGS. THIS TABLE NOTES THE ISSUE, ITS FREQUENCY, AND WHICH CONDITION HAS THE ADVANTAGE IN CONVEYING THAT FEATURE TO ANNOTATORS.

would provide comparable performance to custom-generated 3D images while incurring only a fraction of the data preparation effort. The 2D-L data preparation is "per-map" whereas the 3D is "per-pose", enhancing scalability. Another benefit of the 2D-L map is that it does not require any specialized software to prepare. The 3D approach required custom software to rotate and render the point clouds and robot parking position indicators. Thus, although the 2D-L approach requires the manual annotation of each map, it is still more efficient and easier to scale than the 3D approach.

It is likely that the 3D view provides additional value beyond identification of clutter. For example, it may help people better judge scale in an intuitive way, e.g. judging the width of a hallway, although we did not focus on this point in the current study.

Overall, our results support our hypothesis that the 2D-L maps provided additional context about the potential robot parking locations in the home and thus annotator ratings were more similar to those that were produced from viewing actual 3D depictions of the home environment.

### A. Future Work

This work sets the foundation for a methodology for collecting subjective preferential data about spatial locations at scale, which is necessary for building a product, and which could be applied not only to homes but also to businesses and public or commercial spaces. In this paper we focused on the problem of judging whether a parking location is socially acceptable, but other subjective judgments could also be collected (e.g., whether a stopping location is visible, whether it seems accessible, whether it would be annoying/obtrusive, or even whether it would be aesthetically pleasing). Going beyond potential parking locations, the same kinds of map representations could be used for annotating activity areas, traffic paths, room boundaries and labels, or other information relevant to providing robots with a functional model of human behavior or human preferences.

We used navigational maps obtained by an Astro robot and a Lidar scanner to generate our data, but alternative methods could be leveraged to achieve the same result. For example, some robots may be able to recognize more features in the environment, which would make the 2D map labeling easier. Some robots can generate their own 3D map, in which case an off-board scanning device is unnecessary. In such cases, there is no longer an increased cost of data collection or time to register the maps. However, the challenges of choosing appropriate camera angles for each potential parking position to be annotated, and of fitting features into a single camera field of view, remain. Therefore, our findings will still apply for robots capable of 3D mapping.

A more efficient approach to collecting preferential data about spatial locations could be to have annotators color in regions of parking acceptability. The approach of coloring in regions was originally not considered because it is not feasible with the 3D representations, but based on our finding that a 2D-L map can provide similar results, we could consider a map-coloring approach. It is unclear whether such an approach would sacrifice accuracy, but it seems likely to save a great deal of time, which is another possible advantage of the 2D-L approach.

### B. A Note on HRI Research in Industry

The work reported in this paper was done in the context of a consumer robot product (Astro) at a large company (Amazon), rather than at an academic lab. This introduced a number of constraints that shaped the work. The first constraint, common to all industrial settings, is that the goal is to improve the product, not to do basic science. This has the effect of framing and constraining the problems that industrial HRI researchers work on.

Additional constraints are imposed by the need to maintain company intellectual property and trade secrets. This limits

what we can discuss in terms of implications of the findings for downstream algorithms which rely on this data, and it also limits what can be discussed in terms of future work, as potential future product features are a trade secret.

In line with company policies for user data privacy protection, the annotations for the work were done by an internal annotation team rather than actual users of the robot. This almost certainly affected the annotations that they made, since they did not have the lived experience of being with the robot in their own homes. In an industrial setting, a typical in-home study is hard to execute and publish because of user data privacy controls; we cannot share home maps between users to compare how different people annotate them.

An industrial setting also offers potential benefits for HRI research. Consumer devices are deployed at a scale, often measured in the millions of units, that is impossible in an academic study. If we can design studies that take advantage of this scale, we can do in-the-wild experiments at a scale previously unimaginable. However, these studies will have to be carefully designed to respect user data and privacy, to be transparent to the user, and point towards improving the customer experience.

## VI. Conclusion

A social robot has the potential to assist with many tasks in the home, but it will need to understand and model human spatial preference to effectively position itself in such a complex environment. We propose that crowdsourcing human judgements of spatial preference is an effective technique that can be used to support the development and validation of robot behaviors that rely on such knowledge. Effective crowdsourcing relies on an accurate depiction of the target environment. While 2D navigational maps of the home are easy for us to access and generate, they lack important contextual information needed for annotators to provide useful ratings. 3D data obtained from RGB point-clouds provide a rich and contextual view of the home environment, but the current process to generate such data is both time-consuming and difficult to scale. Labeled 2D maps proved to be a contextually informative medium between the original two approaches. In conclusion, we recommend the labeled 2D map approach as a scalable data presentation method for annotation tasks which require use of human intuition to reason over the social use of space. The findings of this study will enable us to greatly increase the size of our data corpus and improve our annotation process for developing and validating robot features which require spatial understanding in the home.

## ACKNOWLEDGMENT

We thank Hanxiao Fu for his support during this project and acknowledge partial support provided by a research gift from Amazon Devices and Services.

## References

[1] M. M. de Graaf, S. Ben Allouch, and J. A. Van Dijk, "Why would i use this in my home? a model of domestic social robot acceptance," *Human–Computer Interaction*, vol. 34, no. 2, pp. 115–173, 2019.

[2] U. A. Saari, A. Tossavainen, K. Kaipainen, and S. J. Mäkinen, "Exploring factors influencing the acceptance of social robots among early adopters and mass market representatives," *Robotics and Autonomous Systems*, vol. 151, p. 104033, 2022.

[3] "Amazon Astro, household robot for home monitoring," https://www.amazon.com/Introducing-Amazon-Astro/dp/B078NSDFSB, 2021, [Online; accessed 15-February-2023].

[4] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality control in crowdsourcing systems: Issues and directions," *IEEE Internet Computing*, vol. 17, no. 2, pp. 76–81, 2013.

[5] C. A. Cifuentes, M. J. Pinto, N. Céspedes, and M. Múnera, "Social robots in therapy and care," *Current Robotics Reports*, vol. 1, pp. 59–74, 2020.

[6] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, pp. 291–308, 2013.

[7] A. Henschel, G. Laban, and E. S. Cross, "What makes a robot social? a review of social robots from science fiction to a home or hospital near you," *Current Robotics Reports*, vol. 2, pp. 9–19, 2021.

[8] M. Hans, B. Graf, and R. Schraft, "Robotic home assistant care-o-bot: Past-present-future," in *Proc. 11th IEEE international workshop on robot and human interactive communication*. IEEE, 2002, pp. 380–385.

[9] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry, "What is a robot companion-friend, assistant or butler?" in *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2005, pp. 1192–1197.

[10] B. Cagiltay, H.-R. Ho, J. E. Michaelis, and B. Mutlu, "Investigating family perceptions and design preferences for an in-home robot," in *Proceedings of the interaction design and children conference*, 2020, pp. 229–242.

[11] F. G. Prattico and F. Lamberti, "Mixed-reality robotic games: design guidelines for effective entertainment with consumer robots," *IEEE Consumer Electronics Magazine*, vol. 10, no. 1, pp. 6–16, 2020.

[12] M. M. A. de Graaf, "Living with robots: investigating the user acceptance of social robots in domestic environments," 2015.

[13] C. Zhang, L. Zhou, Y. Li, and Y. Fan, "A dynamic path planning method for social robots in the home environment," *Electronics*, vol. 9, no. 7, p. 1173, 2020.

[14] I. B. d. A. Santos and R. A. Romero, "A deep reinforcement learning approach with visual semantic navigation with memory for mobile robots in indoor home context," *Journal of Intelligent & Robotic Systems*, vol. 104, no. 3, p. 40, 2022.

[15] J. Vroon, M. Joosse, M. Lohse, J. Kolkmeier, J. Kim, K. Truong, G. Englebienne, D. Heylen, and V. Evers, "Dynamics of social positioning patterns in group-robot interactions," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2015, pp. 394–399.

[16] J. Vroon, G. Englebienne, and V. Evers, "Detecting perceived appropriateness of a robot's social positioning behavior from non-verbal cues," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 2019, pp. 216–225.

[17] A. Fallatah, B. Stoddard, M. Burnett, and H. Knight, "Towards user-centric robot furniture arrangement," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 2021, pp. 1066–1073.

[18] J. Nauta, C. Mahieu, C. Michiels, F. Ongenae, F. De Backere, F. De Turck, Y. Khaluf, and P. Simoens, "Pro-active positioning of a social robot intervening upon behavioral disturbances of persons with dementia in a smart nursing home," *Cognitive Systems Research*, vol. 57, pp. 160–174, 2019.

[19] M. Shiomi, T. Kanda, D. F. Glas, S. Satake, H. Ishiguro, and N. Hagita, "Field trial of networked social robots in a shopping mall," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2846–2853.

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

[21] J. Ye and A. Sen, "How does Astro localize itself in an ever-changing home?" 2022, accessed on March 7, 2023. [Online]. Available: https://www.amazon.science/blog/how-does-astro-localize-itself-in-an-ever-changing-home