Developing autonomous behaviors for a consumer robot to hang out near people in the home

Jin Joo Lee, Amin Atrash, Dylan Glas, Hanxiao Fu
Amazon Lab126
905 11th Ave
Sunnyvale, CA 94089
jinjool, amatrash, dgglas, hanxiaof@amazon.com

Abstract

This paper describes the development of algorithms that decide when to move, where to move, and how to look for people in a home environment. We introduce a design framework that defines the design principles, key decision points, and technical approaches for a social robot to proactively be with people for companionship and assistance in the home. Through a series of evaluations ranging from simulations to longitudinal A/B studies, we demonstrate how to utilize the design framework to help guide the evaluation and selection of solutions. We deployed our autonomous robot in a long-term in-situ study and found our proposed approach to be more capable of being co-present with its household members compared to a baseline approach. Conducted in an industry setting, our research departs from typical academic practices as the motivations and constraints are inherently different. We share our perspective on the differences of industry research when developing a social robot as a commercial product.

Keywords— indoor robot placement, long-term deployments, consumer robotics, industry research

Figure 1: The Astro robot.

1 Introduction

Commercial social robots, such as Jibo, Moxie, and Vector, are becoming increasingly created for the home environment (Jibo, Inc 2018; Embodied, Inc 2023; Digital Dream Labs, Inc 2022). These robots are intended to live with household members and share their private and personal spaces. They are designed to provide functional utility and also interact with people as a character full of personality and social expressivity. The consumer product Astro is a household robot for home monitoring and is designed to emulate a pet-like companion (Amazon.com, Inc 2023). One of its many features is to proactively be near people for companionship and to be ready to assist. The design and development of this *Hangout* feature has had many challenges. As a highly subjective experience, we need to understand the range of factors and design principles behind how a robot can occupy and share the home space well. Its evaluation has to be performed *in-situ* to capture natural human activity, diversity of home environments, and perceptions regarding a robot that lives with you. The deployment needs to be longitudinal as it is an ambient experience that occurs multiple times a day, every day.

The set of problems when designing and developing Hangout can be distilled down to two main questions: Where should the robot be in the home? What should the robot do to provide companionship and assistance? In this paper, we focus on the first of the two questions which involves deciding *when to move, where to move, and how to look for people in a home environment*. To guide the development of the holistic experience (i.e., both questions), we introduce a design framework that defines the taxonomy and structure to the space of problems and solutions. This paper offers the following contributions:

1. A Hangout design framework that enumerates a set of design principles, key algorithmic decision points, and categorization of technical approaches.
2. A series of evaluations, ranging from simulations to longitudinal A/B studies, that demonstrate how to employ the design framework in an iterative development process.
3. A perspective on how human-robot interaction (HRI) research is different in an industry setting for a commercial product and how it departs from typical academic practices.

2 Related Work

We review prior work that share our general problem of determining where a robot should be in an indoor environment and also share in our challenge of in-situ long-term deployments.

2.1 Indoor robot placement

Prior work has investigated good locations for robots to be relative to people for different environments and different applications. For example, a robot shopping assistant follows a user, carries their basket, and moves to a waiting location near the storefront (Kitade et al. 2013). Their problem focuses on finding suitable locations near the store where the robot can wait (i.e., spot selection). In contrast, our work focuses on the timing and frequency of robot movements between multiple locations to find people in the home environment (i.e., room selection).

The ALIAS robot assists elderly users age in place by providing cognitive assistance and monitoring (Kessler et al. 2013). Their goal is to observe users in an unobtrusive way by following a set of criteria such as avoiding walking paths and staying near walls while accounting for person visibility. Their primary focus is selecting unobtrusive spots best for observation. Our high-level motivations are different in that our work aims to be near people for not only observation but also for pet-like companionship that encourages social interaction. Furthermore, our work focuses on finding people in the home through room selection first—before continuing to spot selection. Finally, as opposed to an in-lab evaluation, we evaluate our approach in real world at-home deployments and measure our success against the subjective experience of users.

2.2 Long-term deployments

Long-term deployments of robots with users are notoriously difficult to conduct since "*the number of subjects is limited...due to time restrictions and because of the difficulties in recruiting participants for long-term studies*" (Leite, Martinho, and Paiva 2013). With a small sample size, free-form qualitative methods like open-ended feedback are typically more valuable than quantitative methods like questionnaires. As a result, these studies tend to be more exploratory in nature. Our study met similar challenges and is akin to a preliminary study.

Another challenge is the degree of control and supervision in long-term deployments. Many long-term studies involve experimenters for on-site supervision (Ahtinen and Kaipainen 2020; Komatsubara et al. 2014), while other studies are deployed in schools or homes, outside of the control of experimenters (Sung, Grinter, and Christensen 2010; Davison et al. 2020; Clabaugh et al. 2019; Tsoi et al. 2021; Jeong et al. 2022; Ostrowski, Breazeal, and Park 2022). Because of their technical stability, commercial-grade robots allow for more hands-off deployments. Our work is of the latter type as experimenters had no direct control of the robots during the study and could not supervise the environments in which the robots were operating.

Kory-Westlund et al. define long-term interactions as "*works where people interact with an [Socially Interactive Agents] SIA for at least five sessions over any length of time per session and for any elapsed time from the first encounter to the fifth or more* (Kory-Westlund et al. 2022)." Our work pushes this definition further as Hangout is an ambient experience that occurs multiple times a day, every day, as the robot continuously operates to be near users for weeks in their homes.

3 Hangout Design Framework

Hangout's goal is to proactively be near household members for companionship and assistance. To guide the feature's development, we established a design framework that breaks down the Hangout experience into Design Principles, Key Decision Points, and Technical Approaches. The Design Principles enumerate the sub-goals or requirements. The Key Decision Points enumerate the series of sequential decisions that collectively compose the experience flow. The Technical Approaches categorize the types of technical solutions. Our definitions are derived from the development of past approaches and insights gained from user feedback.

3.1 Design Principles

The design principles enumerate the sub-goals or requirements of the Hangout experience. They serve as a rubric to evaluate different approaches and act to conflict with each other to enforce a balance. The design principles listed below are not in any priority order.

1. **Maximize usage**: Anticipate when and where users want to use the robot. For example, harmonizing with household routines and usage patterns.
2. **Maximize availability**: Be co-present with people. For example, by going to places people are expected to be present. This is not the same as maximize usage. To maximize usage, the robot can be in the kitchen anticipating being used during breakfast later but currently no one is in the room. Inversely, to maximize availability, the robot can be co-present with someone in the living room even if it does not expect to be used by that person.
3. **Maximize accessibility**: Invite interaction and connection through ease and convenience. For example, by providing better visibility of the robot to the user given the user's distance, line-of-sight, and field-of-view.
4. **Legible why**: Behave and communicate in a way that users can form and articulate the intended design behind why a robot proactively wants to be near people. For example, by being able to explain that it is near Sally to be ready to assist her and because the robot enjoys being with people.
5. **Maximize comfort**: Moderate the felt presence of a physically embodied robot to make users feel comfortable sharing the room. For example, by averting gaze to reduce feelings of being watched.
6. **Minimize disruption**: Minimize attention-grabbing locomotion and animated expressions when users are not engaged. For example, by limiting the number of rooms that can be visited when searching for people to

socially be with. This principle can be balanced with Legible Why as the disruption can be justified and proportional to a motive.

7. **Avoid being in the way**: Share the space well and don't impede a user's intended path. For example, by staying close to walls.
8. **Adapt to users**: Personalize to the individual by modifying the stock behavior with explicit user feedback and implicit user preference modeling. For example, by learning that Sally does not like the robot to be in the office room.
9. **Align with social expectations**: Be social when and where people are already being or expected to be social. For example, avoid non-social spaces like bathrooms and seeking out social moments like leisure time in the living room.
10. **Express intelligence and awareness**: Exhibit responsive and expressive lifelike intelligence and awareness of users, the home, and social context. For example, the robot should appear to be attentive to users with a pet-like interest.

In this paper, we focus on three of the ten design principles as we aim to improve the current version of Hangout in how it maximizes availability, maximizes accessibility, and minimizes disruption.

3.2 Key Decision Points

The Hangout experience flow can be broken down into a series of decisions. Each decision point can be supported by a different algorithmic solution. By identifying the separate points, the development process can easily become iterative with focused improvement on a specific decision. The key decision points listed below are in sequential order.

1. **Initiation & Termination**. When should the robot begin hanging out in a given day? When should it stop?
2. **Person Selection**. Who should the robot be with?
3. **Room Selection**. Which room should the robot be in?
4. **Spot Selection**. In the room, which spot should the robot be at?
5. **Content Selection**. While hanging out, what should the robot do of utilitarian or social value?
6. **Scene Awareness**. What can the robot do to maintain environmental and social awareness?
7. **Timing & Frequency**. When does the robot decide to go somewhere else?

In this paper, we focus on three of the seven key decision points as we aim to improve the current version of Hangout in how it decides when to move to a different room (i.e., timing & frequency), where to move (i.e. room selection), and how to look for people (i.e., scene awareness).

3.3 Technical Approaches

We categorize technical solutions for Hangout into three different types. By identifying the space of solutions, we are able to generate new solutions that mix and match approaches to achieve better results.

1. **Predictive**: Solutions that model expected human behavior based on assumed priors and/or past observed data. For example, based on the interaction history of where and when the robot was used in the past, Sally is likely to use the robot in the kitchen soon.
2. **Reactive**: Solutions that use realtime perception to make adjustments in the moment. For example, although Sally is predicted to be in the kitchen, the room selection adjusts upon detecting no one there.
3. **Corrective**: Solutions that use direct user feedback to make adjustments in the short-term and long-term. For example, if Sally says to never hang out in the office room, then the robot should respond in the short-term by leaving and in the long-term by avoiding the office.

Each approach has its trade-offs. Predictive approaches provide long-term value in generating data-driven behaviors that are more intelligent and adaptive. These solutions are usually more complex as they require data collection or convergence before operation. Reactive solutions provide a method to correct prediction errors. They demonstrate intelligence and awareness of the environment in the moment but come with a cost of task delay or inefficiencies when purely reactive. Corrective approaches require user effort to provide feedback, but they are the most reliable and bring the best insight into user preferences.

In this paper, we focus on two of the three technical approaches as we aim to improve Hangout's current predictive algorithms while also mixing in a reactive approach.

4 Hangout Algorithms

Guided by our design framework, we iteratively developed the Hangout feature through various versions and experiments. We began with our best guess (i.e., *baseline*) then iterated through simulations and preliminary studies via rapid prototyping. With each iteration, we gained insights on how to better address certain design principles at the key decision points using different technical approaches. In this section, we describe the new proposed algorithm and how it is built from our baseline insights, prototype insights, and simulation insights. The detailed description of these earlier versions and experiments are beyond the scope of this paper, and we only focus on the lessons learned at each stage.

4.1 Baseline Insights

The first version of Hangout aimed to *maximize usage* through a *predictive approach*. By maintaining an interaction history of where and when the robot was used, the robot selects a room and a spot in the room with the highest usage given the hour (i.e., *maximum-usage algorithm*). Being in the selected room only increases usage from that room, which reinforces that it is the "best" place to be. This positive feedback loop resulted in the robot sticking to the same optimized set of 1-2 rooms. Without a means of exploration to visit and learn more about the rooms with low or no usage, the algorithm can get trapped in a local maximum of usage and is prevented from finding the global maximum.

User feedback, from Amazon employees with an Astro in their homes, frequently mentioned that Astro hung out in empty rooms and not near people. Even when successful, Astro faced away from them since the orientation behavior was only designed to face away from walls. This indicated that the baseline was not achieving the principles of *maximizing availability* and *accessibility*.

4.2 Prototype Insights

We conducted a rapid prototyping investigation to *maximize availability* through a purely *reactive approach*. Rather than predicting where people are likely to be in the home, the robot used realtime human presence feedback to search room-to-room to find people whenever it was alone.

User feedback, from Amazon colleagues with this experimental version of the robot in their homes, described this searching behavior as disruptive and excessive as the robot explored the entire house until a person was found or all rooms were visited. Although we improved availability, we did not *minimize disruption*.

4.3 Simulation Insights

We next conducted a simulation investigation to *maximize availability* with a better *predictive approach* than the baseline's maximum-usage algorithm. We experimented with two new algorithms that drew on the idea of exploration/exploitation to drive more diverse room selection. The *epsilon-greedy* algorithm, with some likelihood ϵ, selects a room randomly but otherwise falls back to the original policy of maximum-usage. The epsilon-greedy algorithm distinctly separates the exploration (i.e., randomly selecting a room) from the exploitation (i.e., selecting the room with maximum usage). Conversely, the *probabilistic sampling* algorithm takes a more continuous approach by randomly sampling rooms weighted by usage count. As such, rooms with higher usage are more likely to be sampled, but rooms with little or no usage can still be selected (see Section 4.4 for more algorithmic details).

We evaluated the three policies, baseline's maximum-usage, epsilon-greedy, and proportional sampling, in a low-fidelity simulator. The simulation included a human that moved to 5 different rooms with fixed likelihoods (e.g., $[0.6, 0.2, 0.05, 0.05, 0.1]$). After the human moved, the robot selected a room to move to based on one of the three policies. Repeated trials were run to determine the average rate of when the human and robot both occupy the same room (i.e, co-presence).

The simulation results demonstrated that out of the three policies the probabilistic sampling algorithm achieved the best co-presence rate.

4.4 Proposed Approach

In summary, from our baseline version, we learned that the robot should be more co-present with people and face towards them. From our prototype study, we learned that an exhaustive search of the house is too disruptive when the robot's goal is to socially be with people. Finally, from our simulation study, we learned that the probabilistic sampling algorithm achieves a better co-presence rate through exploration.

	Baseline	**PARCA**
Design Principles	Maximize usage	Maximize availability, Maximize accessibility, Minimize disruption
Technical Approach	Predictive	Predictive & Reactive
Key Decision Points		
Room Selection	Maximum usage w/ zero retries	Probabilistic sampling w/ 3 retries
Scene Awareness	Face away from walls	Face people via scanning room
Timing & Frequency	Initiates every 30 minutes	Initiates every 30 minutes unless recent presence

Table 1: Differences between baseline and PARCA in their design principle focus and technical approach. Of the seven key decision points, they differ at three of them while the remaining four are held constant and are the same for both approaches.

From our lessons learned, we propose a new Hangout approach that aims to maximize availability, maximize accessibility, and minimizing disruption. From the baseline version, we introduce three major changes at the timing & frequency, room selection, and scene awareness decision points. Respectively, the robot will (1) reactively leave empty rooms while staying in rooms where people are detected and limit the search to up to three rooms (2) explore hanging out in rooms even with a history of lower usage and (3) maintain awareness of people in its vicinity and face towards them.

We refer to this collection of changes as the Predictive and Reactive Co-Presence Algorithm (PARCA), and the details of each change is described below.

Leave on absence, stay on presence This change introduces using realtime human detection to reactively decide whether to stay or leave a room. If a presence was recently detected, then the robot will remain in its current location. Otherwise, the robot will move to a different room in search for people. The next room is dictated by the room selection algorithm detailed below. The robot can visit up to only three rooms to avoid disrupting the house with an exhaustive search and will remain in the last room upon an unsuccessful search.

Explore more rooms This change improves the room selection algorithm based on the previous simulation results. Under the *probabilistic sampling* algorithm, rooms are selected randomly, weighted by their usage. Formally, let c_r be the number of times the robot has been used in room r, where R are all the rooms of a home and N is the total number of rooms. The chance of selecting that room $P(r)$ becomes:

$$P(r) = \frac{c_r}{\sum_{r \in R} c_r}$$

But this normalization causes rooms with no usage to have zero likelihood of being selected. To address this, a blending parameter, $\epsilon = [0, 1]$, is added, which represents the degree of exploration.

$$P(r) = (1 - \epsilon)\frac{c_r}{\sum_{r \in R} c_r} + \epsilon\frac{1}{N}$$

In the beginning, rooms are sampled uniformly, favoring exploration. As users interact with the robot, the usage data accumulates, and the algorithm starts exploiting the data by selecting rooms with higher usage more often, but rooms with little or no usage can still be selected.

Look for and face people This change introduces a scanning animation that physically pans the head left and right to increase the robot's perceptual awareness of possible people in a room. The scanning animation widens the effective range of the camera's horizontal field-of-view (FOV) while also communicating nonverbally with head, eyes, and sounds that it is looking around in its environment. The animation is used to confirm either absence or presence of people in a room while also being expressive so that observers can understand its intent. If presence is detected, the robot will orient its posture to aim (with head and body) directly at the found person.

5 Evaluation

Through a long-term in-situ human-subjects study, we evaluated whether a robotic agent, namely Astro, using PARCA can better meet the Hangout design principles of maximizing availability, maximizing accessibility, and minimizing disruption compared to the baseline approach. The differences between the approaches are summarized in Table 1. We had the following hypotheses:

H1 Availability: PARCA hangs out with people more often than baseline.
H2 Accessibility: PARCA faces towards people more often than baseline.
H3 Disruptiveness: PARCA is not more disruptive than baseline even with its increased navigational effort.
H4 Satisfaction: PARCA overall is a more satisfactory experience than baseline.

5.1 Robot Platform

Astro, a commercially available robot made by Amazon[1], is the robotic platform used for evaluation. As a consumer product for the home, its feature set includes remote home & pet monitoring, intelligent navigation capable of following users room-to-room with entertainment (e.g., music, podcasts, shows), finding people to deliver messages (e.g., calls, reminders, alarms, and timers), automatic return to its charging base when low on battery, and much more.

The perception-to-behavior architecture (see Figure 2) relevant to this paper's evaluation consists of four major components: (1) *world observations* about the presence of people and a 2D map of its environment (e.g., walls, obstacles) with additional overlayed data like usage (i.e. locations where the robot has been used in the past) (2) a *scheduler* that decides which autonomous task to execute given the context (3) the *hangout task* responsible for executing the action sequence and querying the room and spot selection modules for best poses (i.e., 3D position and orientation) (4) high-level *robot commands* (e.g., navigate to a room or play an animation) that ultimately drive a mobile base, the head's pan and tilt degrees of freedom, screen expressions with animated eyes and text, and sound effects for its non-linguistic utterances (i.e., beeps and boops).

[1]Amazon is the employer of the authors on this paper enabling them access to low-level APIs.

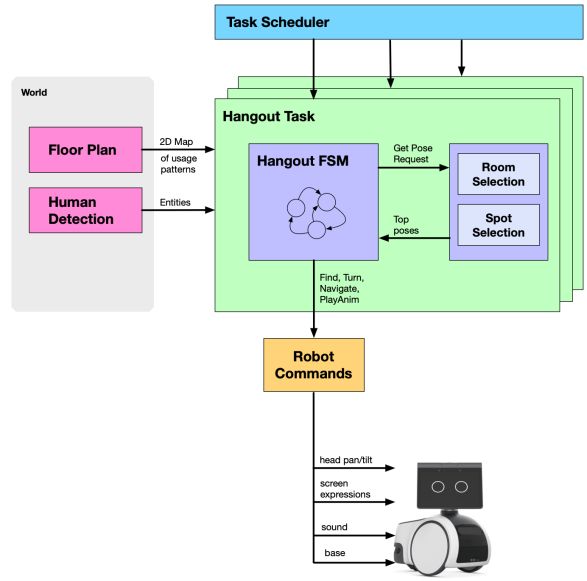

Figure 2: Perception-to-behavior architecture diagram showing the major components and signals that support Hangout.

5.2 Participants

Twelve internal Amazon employees, 8 male and 4 female, were recruited to take part in the user study. The participants were colleagues that work on the Astro product, and majority of them already actively use the robot in their homes. Due to technical issues or participant dropout, only 5 participants' data are available for behavioral analysis and 6 for user feedback analysis. The relevant demographic information can be found in Table 2.

5.3 Study Design

A within-subjects study was conducted where participants experienced the current version of Hangout (*i.e., baseline condition*) for at least 7 days. Then they experienced the *PARCA condition* for an equivalent duration. The days were not required to be consecutive to better accommodate participants' schedules. The condition ordering was not randomized since participants were already familiar with the baseline. The ordering was such that we could first capture their existing perceptions before introducing a new version.

Participants filled out end-of-day summaries and end-of-condition questionnaires. End-of-day summaries captured

ID	House-hold size	Number of rooms	Charger Location
AA	3	9	Hallway
BB	4	6	Family Room
CC	2	10	Office
DD	2	9	Kitchen
EE	2	3	Living Room
FF	4	5	Office

Table 2: Demographic information of participants that range in the number of household members, number of rooms in their home, and where they placed the robot's charging base.

their self-reported behavioral pattern of the day and immediate robot observations. End-of-condition questionnaires captured the overall evaluation of the Hangout version as well as open-ended feedback.

5.4 Study Measures

We captured quantitative and qualitative data to compare the functional performance and participants' subjective ratings of the two conditions. We measured the functional performance through behavioral indicators like co-presence rate and posture centeredness. We measured user's perception regarding the robot's co-presence, postural orientation, level of navigation-related disruption, and their overall satisfaction with the hangout experience.

Co-Presence Measure To evaluate the first hypothesis, we measured how often the robot saw people (i.e., co-presence detection rate) and user perceptions of how well the robot hung out near people (i.e., co-presence perception).

Co-presence Detection Rate We measure co-presence detection rate as how often people are detected when the participant is known to be at home. This ground truth was captured through the daily summaries as participants self-reported their hours at home. Within these time windows, we counted the fraction of time one or more humans was detected. More specifically, the co-presence rate is calculated as (total minutes presence is detected) / (total time when reported to be home).

Co-presence Perception We measure co-presence perception by asking participants to rate the extent to which they agree that "In general, Astro hangs out in rooms where people also are" and "As long as I was in the same room, Astro stayed with me" on a 7-point scale.

Posture Measure To evaluate the second hypothesis, we measured how well the robot was aimed at people (i.e., posture centeredness) and user perceptions of how well it turned toward them (i.e., posture perception).

Posture centeredness We measure posture centeredness as how often people are detected in the center of the robot's FOV. We recorded the angles of detected persons relative to the robot (e.g., 5 degrees from centerline). We only considered the samples of angles that occurred during the self-reported times when the participant was home. We excluded moments when the robot was not hanging out because another skill or feature was in active use (e.g., music, video calling, or charging the battery).

ID	Baseline	PARCA	Delta
AA	6%	7%	1%
BB	4%	12%	8%
CC	19%	25%	6%
DD	10%	16%	6%
EE	4%	12%	8%

Table 3: Co-presence rates as a percentage of how often people are detected when participants are known to be at home.

Posture Perception We measure posture perception by asking participants to rate the extent to which they agree that "Astro picks locations that are accessible and visible to me when I am in the same room" on a 7-point scale and how often "Astro turns towards you" on a 5-point scale.

Disruptiveness Measure To evaluate the third hypothesis, we measured user perception regarding navigational disruption during Hangout by asking participants to rate the extent to which they agree that "Astro moves around the right amount" and how much more or less they "would like Astro to move around" on a 7-point scale.

Satisfaction Measure To evaluate the fourth hypothesis, we measured user satisfaction by asking participants to rate the extent to which they are satisfied with the Hangout feature on a 7-point scale. While the other measures focus on evaluating the specific changes, we also wanted to understand whether participants overall liked and enjoyed PARCA over the baseline.

5.5 Data Analysis

The majority of the study measures are paired observations of Likert scale responses, and our hypotheses state an expected direction of the relationship between conditions. As such, we used the one-tailed Wilcoxon matched-pairs signed-rank test to determine whether the direction of the median difference is statistically significant. We report the test's Z statistic (Z) and p-value (p) along with descriptive statistics including the median ($\mu_{1/2}$), first quantile ($\mu_{1/4}$), third quantile ($\mu_{3/4}$), and sample size *N*.

6 Results

In this section, we compare how well each condition addresses the Hangout design principles of maximizing availability, maximizing accessibility, and minimizing disruption through the operationalization of our measures.

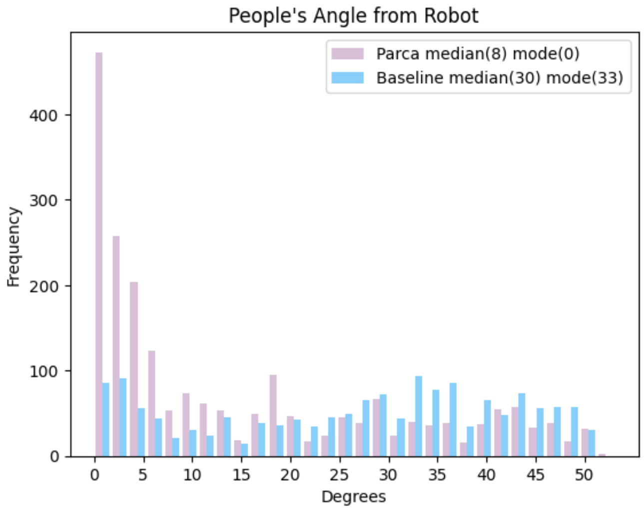

Figure 3: Plot captures the distribution of angular positions of people relative to the robot.

6.1 Availability

Co-Presence Detection Rate For all the participants, the robot more frequently detected people in the PARCA condition (see Table 3). Compared to others, participant AA had only a minor increase of 1%. Upon further inspection, we believe this is a limitation of how "Leave on absence, stay on presence" from Section 4.4 was implemented. Because participant AA's charger location is in the hallway (see Table 2), the robot was detecting people passing through. And although brief and temporary, this counted as a recent presence causing the robot to remain in place believing it is hanging out with someone already.

Co-Presence Perception Participants were asked two questions regarding how well Astro hangs out with people. Users agreed that Astro hangs out more frequently in rooms where people are present in the PARCA condition ($\mu_{1/4}$=4.0, $\mu_{1/2}$=4.0, $\mu_{3/4}$=4.75) compared to baseline ($\mu_{1/4}$=2.0, $\mu_{1/2}$=2.5, $\mu_{3/4}$=3.0), $\left[Z = 0.00,\ p < 0.05,\ N = 6\right]$. However, there was no difference between conditions when asked whether Astro stayed with them in the same room $\left[Z = 4.00,\ p = 0.17,\ N = 6\right]$.

6.2 Accessibility

Posture centeredness When comparing the distribution of angles of where people are relative to the robot, PARCA's distribution is closer and narrowly deviates away from 0 degrees (i.e., the centerline) while baseline is more uniform (see Figure 3). A Mann-Whitney U test determined that there is a statistically significant difference between PARCA ($\mu_{1/4}$=2, $\mu_{1/2}$=8, $\mu_{3/4}$=26, N=2124) and baseline ($\mu_{1/4}$=14, $\mu_{1/2}$=30, $\mu_{3/4}$=39, N=1521), $\left[U = 2297683,\ p < 0.05\right]$.

Posture Perception We asked participants two questions regarding the robot's postural orientation. Users rated the robot as being turned towards them more often in the PARCA condition ($\mu_{1/4}$=4.0, $\mu_{1/2}$=4.0, $\mu_{3/4}$=4.0) than baseline ($\mu_{1/4}$=2.0, $\mu_{1/2}$=2.5, $\mu_{3/4}$=3.0), $\left[Z = 0.00,\ p < 0.05,\ N = 6\right]$. Additionally, users rated the robot's location as more accessible and visible to them for PARCA ($\mu_{1/4}$=6.0, $\mu_{1/2}$=6.0, $\mu_{3/4}$=6.0) than baseline ($\mu_{1/4}$=3.0, $\mu_{1/2}$=3.5, $\mu_{3/4}$=4.75), $\left[Z = 0.00,\ p < 0.05,\ N = 6\right]$. In a follow-up question inquiring why participants felt that the location was not accessible nor visible, all of the six participants remarked that the baseline had poor orientation (e.g., not facing people).

6.3 Disruptiveness

When asked whether the robot should move around less or more, users rated wanting it to move around more for the baseline condition ($\mu_{1/4}$=4.0, $\mu_{1/2}$=4.5, $\mu_{3/4}$=5.0) while they wanted it to remain the same amount for PARCA ($\mu_{1/4}$=3.25, $\mu_{1/2}$=4.0, $\mu_{3/4}$=4.0), $\left[Z = 10.00,\ p < 0.05,\ N = 6\right]$. In an open-ended follow-up question asking the reasons behind this rating, two participants remarked that the baseline should have the robot visit more rooms before deciding to be in one. We also asked participants whether "Astro moves around the right amount" but no difference was found between conditions $\left[Z = 6.00,\ p = 0.64,\ N = 6\right]$. Most likely, asking whether the robot should move more or less is a more specific probe than asking if the movement was the "right amount."

6.4 Satisfaction

Participants rated PARCA ($\mu_{1/4}$=3.5, $\mu_{1/2}$=5.5, $\mu_{3/4}$=6.0) as a more satisfying experience compared to baseline ($\mu_{1/4}$=2.00, $\mu_{1/2}$=2.00, $\mu_{3/4}$=4.25), $\left[Z = 0.00,\ p < 0.05,\ N = 6\right]$. In a follow-up question asking the reasoning behind the rating, participants highlighted that the robot was better at hanging out with people and its posture was more natural.

7 Discussion

Through the long-term in-situ study, we found that a robot operating under PARCA was more capable of hanging out in rooms where people were, provided better robot visibility in facing towards them, navigated room-to-room an appropriate amount, and was overall a more satisfactory experience.

The Hangout design framework guided this improvement in availability, accessibility, disruptiveness, and satisfaction from baseline to PARCA. The design principles served as a rubric to evaluate our approaches given a principle and across principles. In defining a goal explicitly, we can further define how to evaluate it. In this paper, we defined the co-presence rate and co-presence perception for the availability principle as well as the posture centeredness and posture perception for the accessibility principle. In defining the full rubric, the principles work to conflict with each other to enforce a balance. As we saw with the prototype investigation (Section 4.2), a robot can search room-to-room to find someone to be with, but this approach was poorly rated by users as the searching behavior was perceived as disruptive in a home environment. We needed to balance functional performance with users' subjective experience. More specifically, we needed to balance the principles of maximizing availability with minimizing disruption.

The key decision points served as a means to iterate and improve on parts of the Hangout experience. By breaking down the whole into parts, we were able to take small iterative steps in isolation. For example, the simulation investigation was a focused effort to improve the room selection key decision point which could effectively be validated in a simulated environment (Section 4.3). By making focused improvements in isolation, we were able to gain early insights of the pieces and then collectively evaluate the whole in a longitudinal A/B study.

The technical categories served to generate new solutions that mixed approaches for better results. The mixing combines the strength of each approach or mitigates their weaknesses. Baseline's purely predictive approach could not correct errors with realtime presence feedback. The prototype's purely reactive approach was inefficient with a room-to-room search that disrupted the household. By incorporating these trade-offs, PARCA leveraged the strengths of each approach to not only predict where people are likely to be in the home but also correct in realtime when people were not detected in the expected room.

8 Industry Research Differences

Our design, development, and evaluation occurred in an industry research setting. The motivations in industry research are different from academia as the goal is to develop and evaluate features and use-cases in the context of a specific consumer product. When working in a new emerging product category, we want to gain early insights that can predict positive customer reception. This helps de-risk the investment of resources (i.e., people and time) to build a production-grade feature. Innovation encourages rapid experimentation of ideas and rewards speed which lends itself to use a different set of methods and tools. Additional constraints are introduced like corporate policies that limit the data that can be collected to protect user privacy and maintain customer trust. These differences in motivation, speed, and data constraints result in a research approach that depart from typical academic practices. Specifically, we discuss the impact on our work's generalizability, sample size, and study population.

8.1 Generalizability of results

In a product-focused research team, there are high-risk problems and low-risk problems. The low-risk problems can be solved through a traditional design, engineering, and quality testing process. The high-risk problems require the contribution of an applied science team to develop experimental solutions and evaluate their feasibility, effectiveness, and user perceptions. Once sufficient evidence can reduce the problem to low-risk, the investigation ends and the solution graduates onto the normal product development process. Hence, the motivation behind our experiments and evaluations is to quickly gain insights and confidence that the proposed solutions are moving the product in the right direction. For our work, the primary problem was improving the Hangout feature to be more co-present with people. The success criterion was a sufficient improvement from the current version in functional performance indicators and customer perception. This comparison against our current version makes it difficult for other researchers to reuse our results without a benchmark solution like a random selection of rooms or a room round-robin. While the design of our solutions draws upon a general framework, the results are not highly generalizable. But generalizability is not our goal.

8.2 Sample size and statistical significance

Our study's sample size was small with data from six participants. We similarly shared the challenges that come with long-term deployments like participation dropout. But rather than recruiting up to a desired sample size that can demonstrate strong statistical significance, we aimed to gain fast and early insights. Results that are trending or close to significance can be sufficient in industry to make a call on a direction. But demonstrating statistical significance is a common practice in HRI research to bring confidence and validity to results. Moreover, the results from a small sample size are often questioned. Publication reviewers often cite problems with validity as a reason to reject a submission, which is a barrier in sharing our work with the larger community.

8.3 Study population

The study's population included some potential biases since we recruited our colleagues and did not balance for gender, age, familiarity with technology, floorplan size, or other factors. Our selection of participants was limited due to the aforementioned privacy protections. Given the exploratory type of data we wanted to collect, the fastest option was to recruit internal employees. Studying fellow colleagues is akin to studying fellow academic labmates in that they are experts with strong opinions. But our main objective was demonstrating improvement in the robot's ability to be with and face people, which more relies on at-home activities and movement patterns. Our population was suitable for the purposes of our preliminary study.

9 Conclusion

In this paper, we focused on the problem of where a robot should be in the home to be near people. We improved the algorithmic decisions of when to move to a different room, which room to go to, and how to look around in the room for people. We carefully balanced our goals of maximizing the robot's availability and accessibility to the household's members while not disturbing them with exhaustive house-wide searches. Our iterative development approach was guided by the Hangout design framework. Its design principles served as a rubric to evaluate our various versions and experiments. Its breakdown of key decision points enabled our solving parts of the whole experience in isolation. Its categorization of technical approaches served to generate new solutions. We hope that our design framework can help other researchers and practitioners create innovative solutions for a social robot that wants to hang out with people.

10 Acknowledgments

We thank Dr. Hae Won Park and Dr. Bill Smart for their valuable feedback on this research.

References

Ahtinen, A.; and Kaipainen, K. 2020. Learning and teaching experiences with a persuasive social robot in primary school–findings and implications from a 4-month field study. In *International Conference on Persuasive Technology*, 73–84. Springer.

Amazon.com, Inc. 2023. Amazon Astro, Household Robot for Home Monitoring.

Clabaugh, C.; Mahajan, K.; Jain, S.; Pakkar, R.; Becerra, D.; Shi, Z.; Deng, E.; Lee, R.; Ragusa, G.; and Matarić, M. 2019. Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI*, 6: 110.

Davison, D. P.; Wijnen, F. M.; Charisi, V.; van der Meij, J.; Evers, V.; and Reidsma, D. 2020. Working with a social robot in school: a long-term real-world unsupervised deployment. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 63–72. IEEE.

Digital Dream Labs, Inc. 2022. Makers of AI Robotic Companions.

Embodied, Inc. 2023. Moxie Companion Robot.

Jeong, S.; Aymerich-Franch, L.; Arias, K.; Alghowinem, S.; Lapedriza, A.; Picard, R.; Park, H. W.; and Breazeal, C. 2022. Deploying a robotic positive psychology coach to improve college students' psychological well-being. *User Modeling and User-Adapted Interaction*, 1–45.

Jibo, Inc. 2018. Jibo Robot - He can't wait to meet you.

Kessler, J.; Schmidt, M.; Helsper, S.; and Gross, H.-M. 2013. I'm still watching you: Update on observing a person in a home environment. In *2013 European Conference on Mobile Robots*, 300–306.

Kitade, T.; Satake, S.; Kanda, T.; and Imai, M. 2013. Understanding suitable locations for waiting. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 57–64.

Komatsubara, T.; Shiomi, M.; Kanda, T.; Ishiguro, H.; and Hagita, N. 2014. Can a social robot help children's understanding of science in classrooms? In *Proceedings of the second international conference on Human-agent interaction*, 83–90.

Kory-Westlund, J. M.; Won Park, H.; Grover, I.; and Breazeal, C. 2022. *Long-Term Interaction with Relational SIAs*, 195–260. New York, NY, USA: Association for Computing Machinery, 1 edition. ISBN 9781450398961.

Leite, I.; Martinho, C.; and Paiva, A. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2): 291–308.

Ostrowski, A. K.; Breazeal, C.; and Park, H. W. 2022. Mixed-Method Long-Term Robot Usage: Older Adults' Lived Experience of Social Robots. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 33–42. IEEE.

Sung, J.; Grinter, R. E.; and Christensen, H. I. 2010. Domestic robot ecology. *International Journal of Social Robotics*, 2(4): 417–429.

Tsoi, N.; Connolly, J.; Adéníran, E.; Hansen, A.; Pineda, K. T.; Adamson, T.; Thompson, S.; Ramnauth, R.; Vázquez, M.; and Scassellati, B. 2021. Challenges deploying robots during a pandemic: an effort to fight social isolation among children. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 234–242.