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

Jin Joo Lee, Amin Atrash, Dylan F. Glas, Hanxiao Fu[1]

***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 as a tool to guide the development of a social robot to proactively be with people for companionship and assistance in the home. Through a series of experiments 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 approach departs from typical academic practices as the motivations are inherently different. We share our perspective on the differences of industry research when developing a social robot as a commercial product.**

I. Introduction

Commercial social robots, such as Jibo, Moxie, and Vector, are becoming increasingly created for the home environment [1]–[3]. 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 [4]. 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 users' 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.

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 a structure to the space

[1]All authors are with Amazon Lab126 Consumer Robotics, 905 11th Ave, Sunnyvale, CA, USA {jinjool, amatrash, dgglas, hanxiaof}@amazon.com

Fig. 1: The Astro robot.

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 categories of technical approaches.
2) A series of experiments, 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.

II. Related Work

Our work is situated in the general problem of determining where a robot should be in the home to be near people. But we differentiate our work from prior work by highlighting (1) our goal of companionship, (2) our long-term deployment, and (3) our focus on the room selection problem.

A. Mobile Companion Robots

Robots in indoor environments search to find people or monitor them for observation. In simulated environments, robots search to find people to remind them of an upcoming recreational activity or to deliver urgent goods like coffee [5], [6]. In real deployments, robots try to find multiple people in an university building before a deadline or a single person through an apartment-wide search [7], [8]. Lastly, a robot assists elderly users age in place by providing cognitive assistance and monitoring [9]. These prior works focus on the time-constrained person-search problem or the unobtrusive human monitoring problem. In contrast, our primary goal is to provide pet-like companionship that encourages social interaction and engagement with the robot. Our problem does not have a time-bounded urgency nor the strict requirement to always successfully find people. Our constraint is to be co-present with people while also balancing human-factors considerations like minimizing the disruption navigation introduces into the home.

B. 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*" [10]. We met similar challenges in recruitment and retention for our 2-week-long study. Another challenge is the degree of control and supervision. Many long-term studies involve experimenters for on-site supervision [11], [12], while other studies are deployed in schools or homes, outside of the control of experimenters [13]–[18]. 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 and could not supervise the environments in which the robots were operating.

C. Indoor Robot Placement

Prior work has investigated good locations for robots to be relative to people for different environments and different applications. A robot shopping assistant selects suitable locations near the storefront where the robot can wait [19]. A monitoring robot selects unobtrusive spots best for user observation in the home [9]. These works focus on the problem of *spot selection* to determine the best spot within a region for waiting or monitoring based on a set of criteria such as avoiding walking paths, staying near walls, and proxemics. In contrast, our work focuses on the timing and frequency of robot movements between multiple regions to find people in the home environment (i.e., *room selection*).

III. 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 were derived from the development of past approaches and insights gained from user feedback.

A. 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) **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.
6) **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.
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 norms**: 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.

B. 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).

C. 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 in realtime upon detecting the absence of people.
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 algorithm by mixing in a reactive approach.

IV. Early 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., our current V1 version) 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. 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.

A. V1 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*). However, we found that being in the selected room only increases usage for 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 about the rooms with low-or-no usage, the algorithm can get trapped in a local maximum and is prevented from finding the global one.

Overall, customers enjoyed having Astro near them instead of remaining on its charging base. However, they mentioned that Astro sometimes 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 V1 was not achieving the principles of *maximizing availability* and *accessibility*.

B. 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.

Customer feedback 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*.

C. Simulation Insights

We next conducted a simulation investigation to find a *predictive approach* that can *maximize availability* better than V1'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 V-B for details).

We evaluated the three policies—V1'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. If the human and the robot are in the same room, there is a chance of usage occurring.

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.

V. Proposed Hangout Algorithm

In this section, we describe the new proposed algorithm and how it is built from our previous insights. In summary, from our V1 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.

From our lessons learned, we propose a new Hangout approach that aims to maximize availability, maximize accessibility, and minimizing disruption. We introduce three major changes from the V1 version 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 our second version, or V2, and the details of each change is described below.

A. 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.

B. 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.

	V1	**V2**
Design Principles	Maximize usage	Maximize availability, Maximize accessibility, Minimize disruption
Technical Approach	Predictive	Predictive & Reactive
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 I: Differences between V1 and V2 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.

C. 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.

VI. Evaluation

Through a long-term in-situ human-subjects study, we evaluated whether a robotic agent, namely Astro, with the Hangout V2 version can better meet the design principles of maximizing availability, maximizing accessibility, and minimizing disruption compared to V1. The differences between the approaches are summarized in Table I. We had the following hypotheses:

1) **Availability**: V2 hangs out with people more often than V1.
2) **Accessibility**: V2 faces towards people more often than V1.
3) **Disruptiveness**: V2 is not more disruptive than V1.
4) **Satisfaction**: V2 is a more satisfactory experience than V1.

A. Robot Platform

Astro, a commercially available robot made by Amazon, is the robotic platform used for evaluation. As a consumer product for the home, its feature set includes remote home

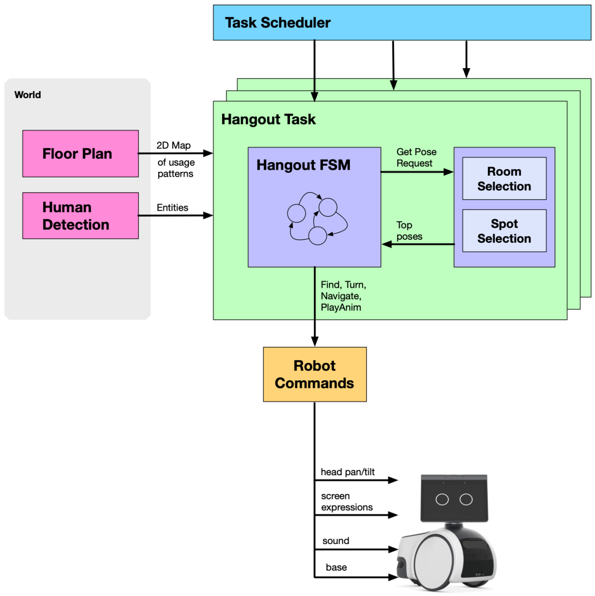

Fig. 2: Perception-to-behavior architecture diagram showing the major components and signals that support Hangout.

ID	Household 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 II: 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.

& 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, head pan and tilt motors, screen expressions with animated eyes and text, and sound effects for its non-linguistic utterances (i.e., beeps & boops).

B. 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 II.

C. Study Design

A within-subjects study was conducted where participants experienced the current version of Hangout (i.e., V1 condition) for at least 7 days. Then they experienced the V2 condition for an equivalent duration. Participants filled out end-of-day summaries and end-of-condition questionnaires. End-of-day summaries captured 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.

D. 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.

1) 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). 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 measured the length of time one or more humans were detected. More specifically, the co-presence rate is calculated as (total minutes presence is detected) / (total time when reported to be home). We measure *co-presence perception* by asking participants to rate the extent to which they agree that "[CP1] In general, Astro hangs out in rooms where people also are" and "[CP2] As long as I was in the same room, Astro stayed with me" on a 7-point scale.

2) 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). 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). We

ID	V1	V2	Delta
AA	6%	7%	1%
BB	4%	12%	8%
CC	19%	25%	6%
DD	10%	16%	6%
EE	4%	12%	8%

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

measure *posture perception* by asking participants to rate the extent to which they agree that "[PO2] 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 "[PO1] Astro turns towards you" on a 5-point scale.

3) 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 "[DS2] Astro moves around the right amount" and how much more or less they "[DS1] would like Astro to move around" on a 7-point scale.

4) Satisfaction Measure: To evaluate the fourth hypothesis, we measured user satisfaction by asking participants to rate "[ST] how satisfied are you with Astro's 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 V2 over V1.

E. Data Analysis

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), p-value (p), and sample size (N) as well as illustrate the descriptive statistics using box and whisker plots in Figure 4.

VII. Results

A. Availability

1) Co-Presence Detection Rate: For all the participants, the robot more frequently detected people in V2 (see Table III). A two proportion z-test showed that there is a statistically significant difference in the proportion of time Astro detected people in V1 (36 hours out of 314) versus V2 (54 hours out of 302), $\left[Z = 2.25, p < 0.05, N = 5\right]$.

2) Co-Presence Perception: Users agreed that Astro hangs out more frequently in rooms where people are present in V2 compared to V1, $\left[Z = 0.00, p < 0.05, N = 6\right]$. However, there was no difference between conditions when asked whether Astro stayed with them in the same room $\left[Z = 4.00, p = 0.17, N = 6\right]$. Most likely, this question was impacted by a different feature that automatically triggers Astro to return to its charging dock when the battery is low.

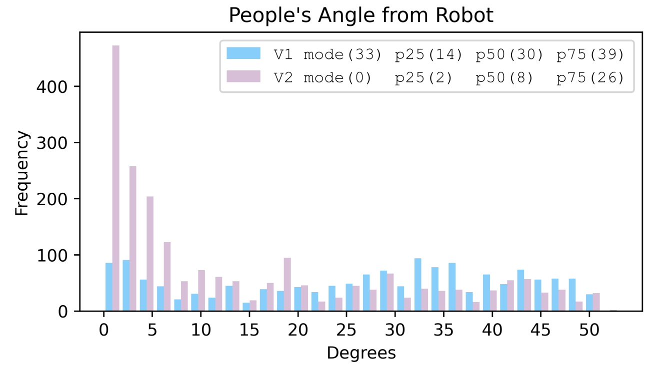

Fig. 3: Plot captures the distribution of angular positions of people relative to the robot. The legend reports the distribution's mode, median (p50), first quantile (p25), and third quantile (p75).

B. Accessibility

1) Posture centeredness: When comparing the distribution of angles of where people are relative to the robot, V2's distribution is closer and narrowly deviates away from 0 degrees (i.e., the centerline) while V1 is more uniform (see Figure 3). A Mann-Whitney U test determined that there is a statistically significant difference, $\left[U = 2297683, p < 0.05\right]$.

2) Posture Perception: Users rated the robot as being turned towards them more often in V2 than V1, $\left[Z = 0.00, p < 0.05, N = 6\right]$. Additionally, users rated the robot's location as more accessible and visible to them for V2, $\left[Z = 0.00, p < 0.05, N = 6\right]$. In a follow-up question inquiring why participants felt that the location was not accessible nor visible, all of the six participants remarked that V1 had poor orientation (e.g., not facing people). In an open-ended question asking "What do you think of Astro's gazing behavior," participants for V1 wanted more interactive and purposeful looks, while for V2 said it is more attentive and organic but also boring and frozen.

C. Disruptiveness

When asked whether the robot should move around more or less, users rated wanting it to move around more for V1 while they wanted it to remain the same amount for V2 even with its increased navigational effort, $\left[Z = 10.00, p < 0.05, N = 6\right]$. In an open-ended follow-up question asking the reasons behind this rating, two participants remarked that V1 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 better probe than asking if the movement was the "right amount."

D. Satisfaction

Users rated V2 as a more satisfying experience compared to V1, $\left[Z = 0.00, p < 0.05, N = 6\right]$. 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.

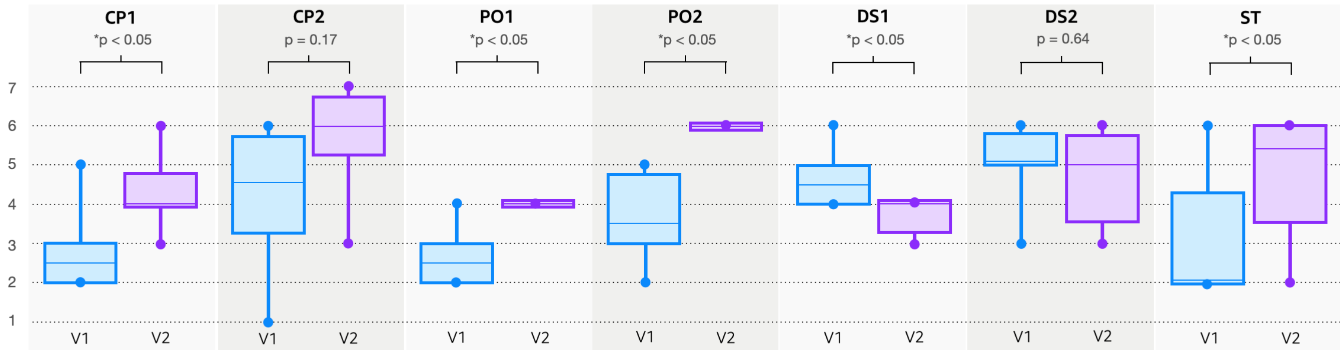

Fig. 4: Plot captures the descriptive statistics of user's perception regarding the robot's co-presence (CP), postural orientation (PO), navigation-related disruption (DS), and overall satisfaction (ST) with the Hangout experience.

VIII. Discussion

Through the long-term in-situ study, we found that our V2 version 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 V1 to V2. 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 detection rate and co-presence perception for the availability principle as well as posture centeredness and posture perception for the accessibility principle. The design principles work to conflict with each other to enforce a balance. As we saw with the prototype investigation (Section IV-B), 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 too disruptive in a home environment. We needed to balance functional performance with users' subjective experience through the balance 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 can take small iterative steps in isolation. For example, the simulation investigation was a focused effort to improve room selection, which could effectively be validated in a sim environment (Section IV-C). 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. V1'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, V2 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.

	Academia	Industry	Differing Practice
Goals	Pursuit of science	High-value product	—
Success Criteria	Generalizability & reuse	CSATs & KPIs	Multiple changes in A/B, No benchmark condition
Result Confidence	Statistical significance	Close or trending	Gain early insights, Small sample size
Population	IRB-protected volunteers	Privacy-protected consumers	Biased population
Continuity	One-shot studies	Continuous versions	No counterbalancing

TABLE IV: Differences between academia and industry that result in differing practices in the study design.

IX. Industry Research Differences

Our design, development, and evaluation occurred in an industry research setting. The motivations in industry are different from academia as the ultimate goal is to create a high-value product that consumers want and need. Because their underlying goals are different, their research approaches also diverge. In this section, we discuss how academia and industry are different in their success criteria, result confidence, population, and continuity. We further highlight how these differences influenced our study design and our decisions to (1) evaluate multiple changes in a single A/B comparison (2) not include a benchmark condition (3) keep a small sample size (4) recruit a biased population (5) not counterbalance the conditions (see summary in Table IV).

A. Success Criteria

Our success criteria was to demonstrate a sufficient improvement from the first version of Hangout in *key performance indicators* (KPIs) like co-presence detection rate and in *customer satisfaction ratings* (CSATs) like users' perception regarding disruptiveness. Since our goal is to improve KPIs and CSATs, we want to include as many changes that can improve these measures. We are not interested in studying the isolated effects of each change as we are solely interested in comparing against our current version. This makes it difficult to reuse our results without a benchmark solution like a random selection of rooms or round-robin the rooms. But the generalizability of results is not our goal.

B. Result Confidence

In a product-focused organization, there are high-risk problems and low-risk problems. The low-risk problems can be solved through a traditional software development lifecycle of design, engineering, and quality testing. The high-risk problems require the contribution of an applied science team to develop experimental solutions and evaluate their feasibility, KPIs, and CSATs. Once sufficient evidence can reduce the problem to low-risk, the investigation ends and the solution graduates onto the traditional software development lifecycle. As such, the researcher's goal is to quickly gain early insights—through prototyping, simulations, and small user studies—that the proposed solutions are moving the product in the right direction.

Our study's sample size was small with a total of six participants. We shared the challenges that come with long-term deployments like participation dropout. But rather than recruiting up to a desired sample size for a strong statistical significance, our aim was to quickly gain 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 to results. Moreover, the results from a small sample size are often questioned.

C. Population

Corporate policies limit the data that can be collected to protect customer's privacy and maintain their trust. Given the exploratory data our study collected, our only option was to recruit colleagues. Studying fellow employees is akin to studying fellow academic labmates. Both are void of balancing for gender, age, familiarity with technology, and other factors. As such, we have to be discerning on the conclusions that can be drawn from this biased population. Our main objective was demonstrating improvement in the robot's ability to be with and face people. Since this more relies on at-home activities and movement patterns, our population was suitable for the purposes of our study.

D. Continuity

One of the benefits working in industry is the continuity of the work. Researchers can continuously improve on features by iterating from customer feedback. Each new version is then released as a software update. Our conditions were not counterbalanced because preserving the ordering emulates our customer's experience. While academics counterbalance to remove potential effects the ordering introduces, we are interested in and subject to those ordering effects.

X. Conclusion

Guided by our design framework, we developed a social robot that was capable of being near people in a home environment for companionship. We hope that our framework can help other researchers and practitioners create innovative solutions for a mobile robot companion. And whether that work is done in academia or industry, the differing practices should not become a barrier in sharing research between the communities. As in the end, we are working towards solving the same problems.

Acknowledgments

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

References

[1] Jibo, Inc. (2018) Jibo robot - he can't wait to meet you. [Online]. Available: https://jibo.com

[2] Embodied, Inc. (2023) Moxie companion robot. [Online]. Available: https://embodied.com

[3] Digital Dream Labs, Inc. (2022) Makers of ai robotic companions. [Online]. Available: https://www.digitaldreamlabs.com

[4] Amazon.com, Inc. (2023) Amazon astro, household robot for home monitoring. [Online]. Available: https://www.amazon.com/astro

[5] M. Schwenk, T. Vaquero, and G. Nejat, "Schedule-based robotic search for multiple residents in a retirement home environment," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2014, p. 2571–2577.

[6] G. D. Tipaldi and K. O. Arras, "I want my coffee hot! Learning to find people under spatio-temporal constraints," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 1217–1222.

[7] S. C. Mohamed, S. Rajaratnam, S. T. Hong, and G. Nejat, "Person finding: An autonomous robot search method for finding multiple dynamic users in human-centered environments," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 433–449, 2020.

[8] M. Volkhardt and H.-M. Gross, "Finding people in home environments with a mobile robot," in *European Conference on Mobile Robots*, 2013, pp. 282–287.

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

[10] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[11] A. Ahtinen and K. Kaipainen, "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*. Springer, 2020, pp. 73–84.

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

[13] J. Sung, R. E. Grinter, and H. I. Christensen, "Domestic robot ecology," *International Journal of Social Robotics*, vol. 2, no. 4, pp. 417–429, 2010.

[14] D. P. Davison, F. M. Wijnen, V. Charisi, J. van der Meij, V. Evers, and D. Reidsma, "Working with a social robot in school: a long-term real-world unsupervised deployment," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 63–72.

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

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

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

[18] A. K. Ostrowski, C. Breazeal, and H. W. Park, "Mixed-method long-term robot usage: Older adults' lived experience of social robots," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2022, pp. 33–42.

[19] T. Kitade, S. Satake, T. Kanda, and M. Imai, "Understanding suitable locations for waiting," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2013, pp. 57–64.