

# Designing for Exit: How to Let Robots Go

Elin Björling<sup>1</sup> and Laurel D. Riek<sup>2</sup>

<sup>1</sup>Department of Human-Centered Design and Engineering, University of Washington, Seattle, USA; email: [bjorling@uw.edu](mailto:bjorling@uw.edu)

<sup>2</sup>Computer Science and Engineering, University of California San Diego, San Diego, USA; email: [lriek@eng.ucsd.edu](mailto:lriek@eng.ucsd.edu)

We Robot 2022. pp. 1–16  
Copyright © 2022 by E. Bjorling and L. Riek.  
All rights reserved

## Keywords

Social Robots, Design, Exit, Robot Ethics, Human-Robot Interaction, Healthcare Robotics, Educational Robotics

## Abstract

The field of human-robot interaction (HRI) has, for decades, focused on sustaining engagement with robots. However, all interactions with robots will end, but this notion is neglected in both the literature and in practice, which has led to deleterious effects for vulnerable users. In this paper, we introduce a new design principle: Designing for Exit, a deliberate design process which leads to, and in many cases assumes, a well-defined ending. Ultimately, the goal is to give users more control over their engagements and disengagements with robots, and give guidelines to robot creators on how to best plan for and facilitate exit. We define and give examples of the types of exit in HRI, and introduce five exit principles for researchers and designers to consider to raise awareness of and ideally mitigate some of the potential harms of exit. We also provide suggestions for how researchers and robot designers can collaboratively build an exit plan with users.

## 1. Introduction

One of the key tenants in human robot interaction (HRI) is engagement, which Sidner et al. [61] describe as how to design interactions so that “participants establish, maintain, and end their perceived connection” with robots. The field of HRI has largely focused on how to establish and maintain connections, but rarely on how to end them. However, thinking about the end of engagement is particularly important as the field moves toward longitudinal robot deployments.

For some populations, such as people with dementia or children, ending the use of a robot in an unintended or unplanned way may lead to deleterious effects [30, 59]. Many cases have been reported both anecdotally and in the literature where robots have been successfully used, but then abruptly removed without a user’s consent or knowledge, adversely affecting their cognitive and mental health. For example, removing the petlike robot Paro from nursing homes has been shown to lead to depression [26], and removing the Huggable and Aflac duck robots from hospitalized children can lead to them being upset [45].

Regardless of a population’s age or cognitive state, the literature suggests most people have a tendency to form deep attachments to robots [16]. They name them, dress them, and, when they stop working, hold funerals for them [12, 18]. These attachments are profound, and can even transcend a robot’s degree of humanlikeness [69, 13, 68].

Prior work explores how and why this is problematic, from the perspective of users being easily susceptible to manipulation by predatory technology companies and other ne’er-do-wellers [23, 30, 54]. Thus, it is important robot designers and researchers alike find ways to design off-ramps from their technology and their users - how, where, and in what ways can we create interactive experiences that will end?

In this paper, we introduce a new design principle: *Designing for Exit*. This serves as a direct counterpoint to “stickiness”, a common industrial design goal intended to sustain engagement for as long as possible in order to maximize profits. Here, we propose a deliberate design process which leads to, and in many cases assumes, a well-defined ending. Ultimately, the goal is to give users more control over their engagements and disengagements with robots, and give guidelines to robot creators on how to best plan for and facilitate exit.

In Section 2, we introduce key terminology, and discuss the concept of exit within related areas of healthcare and design. Then, in Section 3, we discuss various factors which affect how users experience exit, and give examples of exit within HRI. In Section 4, we propose five principles for researchers and designers to consider to support how they can raise awareness of and ideally mitigate some of the potential harms of exit. Then, in Section 5, we give examples of how we, the authors, have applied these principles to design for exit within our own longitudinal HRI deployments. Finally, in Section 6, we give suggestions on how to build an exit plan.

## 2. Background

### 2.1. Human Robot Interaction (HRI): A brief overview

In this paper we will be discussing exit within the context of the field of HRI. It is a broad field that explores how people interact with robots across multiple dimensions, including social and physical interaction, proximate and remote interaction, and interaction across a range of different environments and contexts. For the scope of this paper, we focus on HRI



**Figure 1**

In this paper, we consider HRI in terms of social interaction or physical interaction with robots. **Social Interaction (two images on left):** *Left:* One person and one robot have a social interaction, in the context of robot delivered neurorehabilitation [29]. *Right:* Multiple people interact socially with a robot in a public place [27]. Robots can be non-mobile (left) or mobile (right), and may support a range of interactive modalities. They also may have varying degrees of autonomy. **Physical Interaction (two images on right):** *Left:* One human and one robot interact physically to achieve a goal. There is not necessarily a social component, though it is interactive [36]. *Right:* A mobile robot at a park senses groups of people, but does not necessarily interact with them [66, 65]. Here, the people are more bystanders and or physical interactants rather than direct robot users.

within human social environments (HSE), which is any space where humans and robots are synchronously and proximately interacting [50].

We note that robots in HSEs are not necessarily *sociable robots* [8], in that they are designed and intended for social interaction, but simply that they share a space with humans [50]. Extensive prior work shows that people tend to anthropomorphize nearly anything that they perceive to have agency regardless of how humanlike it appears [46, 32, 24], so when discussing exit in this paper, we consider a range of robot roles, morphologies, and behaviors.

Another important dimension within HRI is level of autonomy. Robots can range from fully teleoperated, where an operator or wizard<sup>1</sup>, controls all aspects of the robot’s behavior, to fully autonomous, where a robot makes all operating decisions itself. In between, there are a range of mixed and shared control scenarios, where some functions of the robot are controlled by artificial intelligence (AI) and some by a human [4, 49, 35].

Regardless of a robot’s level of autonomy, nearly all real world HRI features some kind of “robot wrangler” [64], a human who is responsible for ensuring the robot continues to function, stay charged, and overall ensures a smooth and successful interaction. As we will discuss later, this is a particularly important function to consider when designing for exit, as the function of robot wrangling eventually requires exit, or the job needs to be transferred to an end user.

All HRI is situated within an environment or domain, which shapes the tasks the robots can engage in, the roles and functions they can adopt, and the interaction itself [19]. Robot types include assistive robots, educational robots, service robots, clerical robots, and entertainment robots, which can operate across home and industrial settings, clinical and transportation settings, in workplaces, and others [50]. HRI takes place across multiple time scales, but increasingly the field is moving toward longitudinal, real-world robot de-

<sup>1</sup>Generally if a robot is being remotely embodied by a person in order to engage in a remote space, such as in cases of telemedicine, teleschool, etc, one would describe this person as an *operator*. However, if the remote person is hidden to any co-present humans, one would describe the person as a *wizard*.

ploysments [31, 15, 71, 29].

Both robots and people can adopt a range of different roles and relationships during HRI [19, 57], which can have a strong influence on a user’s experience of exit. Both robots and people can adopt roles including supervisor, operator, mechanic, peer, bystander, and mentor (see [19, 57] for an extensive discussion). These roles can intersect with other parties beyond the primary person interacting with the robot; for example, children will tell secrets to robots they view as peers, but that are actually wizard-of-oz controlled by adults [70, 9], raising ethical questions about information disclosure, and, ultimately, control over both engagement and exit.

Finally, we introduce the following nomenclature to describe human stakeholders in HRI, adopted from Riek [51]:

- **Primary Users:**
  - 1) **Direct Robot Users (DRUs):** A person / persons directly interacting with a robot [51]. DRUs interact with robots across multiple different sectors (manufacturing, education, clinical, home), and also includes participants in research studies.
  - 2) **DRU-related users:** People whom have relationships with DRUs and may direct aspects of interaction, such as caregivers, family members, friends, clinicians, managers, etc. Frequently these individuals serve as gatekeepers for a DRU’s engagement with or exit from technology.
- **Secondary Users:**
  - 1) **DRU-adjacent users / bystanders:** People who share physical space with a DRU but are not related to them or likely to interact with them, such as someone physically co-present in the same space, or walking in the same direction.
  - 2) **Robot makers:** People who “design, build, program, instrument, or research robotics technology” [51]
  - 3) **Robot Customizers/Editors:** People who take an existing robot platform and customize it for their user population.
- **Tertiary Users:**
  - 1) **Administrators:** People who are engaged with the primary users, but not engaged with the robot. For example, a principal at a school or the chief information officer at a hospital.
  - 2) **Beneficiaries:** People who benefit from the relationship between the user and the robot, but are not directly interacting with either type of user. For example, an insurance company may benefit from people using a robot to exercise to lower their cholesterol.

## 2.2. Exit in Healthcare

When designing and deploying health interventions, researchers and clinicians are increasingly turning to the evidence-based practice of de-implementation. De-implementation refers to “reducing or stopping the use or delivery of services or practices that are ineffective, unproven, harmful, overused, or inappropriate” [47]. Generally this takes place to reduce healthcare costs, unnecessary and/or harmful treatments, or replacing an outdated practice with a more effective or new one [47].

De-implementation is a complex process - it can be challenging to act beneficently on both an individual and public health level. For example, an intervention may be withheld from minoritized individuals who experience health disparities due to structural racism,

sexism, classism, gender binarism, etc., which further sows distrust. (Or, conversely, an intervention which is no longer useful or safe continues to be administered). Kerkoff et al. [28] argue that it is important to adopt a health equity, stakeholder-centered lens from the outset of creating an intervention all the way through to de-implementation.

### 2.3. Exit in Industry

As de-implementation is discussed in healthcare, planned obsolescence is quite common in the technology sector. A central dark design pattern for technologies is planned obsolescence, which is “is the production of goods with uneconomically short, useful lives so that customers will have to make repeat purchases” [10]. Matisonn [33] defines several categories of planned obsolescence, including:

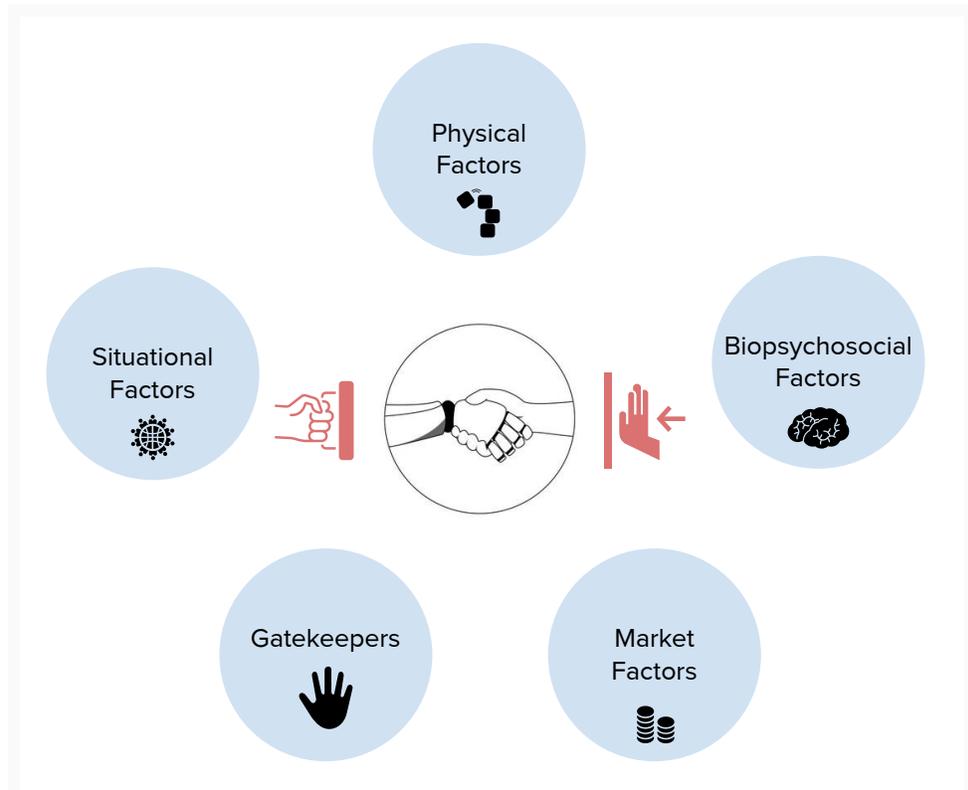
- *Technological obsolescence*, where a change is so significant prior expertise to use the system is obsolete (e.g., typewriters)
- *Limited functional lifespan*, where a manufacturer intentionally “death dates” a product so it will cease to function after a predefined period of time and will need to be replaced (e.g., phone batteries, toothbrushes)
- *Stylistic obsolescence*, where the design of the new product renders the old one aesthetically displeasing, creating “reduced psychological satisfaction” (e.g., an OLED television vs. a CRT) [33, 37]
- *Functional enhancements via new features*, where a product is slightly changed to include new features, and parts / software for the older product is no longer maintained (e.g. a new cell phone model).

Planned obsolescence, by design, leads to a lack of user autonomy. Most planned obsolescence strategies are opaque to users, and they simply must accept that either their product will fail, stop being supported, or will be subjected to manipulation by marketers (sometimes via the robot itself) to upgrade [30, 23, 48]. Additionally, as Maycroft [37] writes, a person’s ability to understand the world and modify objects within it is quickly diminishing. “In a world in which shoes cannot be resoled, transistor radios cannot be repaired, clothes cannot be mended, etc, the opportunities for obsolescence flourish.” Maycroft argues that the only “remedy” for this lost autonomy is over-consumption [37], which in addition to raising a host of broad ethical issues, is harmful to the planet and to those of lower socioecotonomic status [73, 37].

To address these concerns, Matisonn [33] outlines multiple arguments for user-centered design and transparency around planned obsolescence, in order to promote ethical leadership and beneficence on part of companies. Additionally, Mayoral-Vilches et al. [38] introduce the concept of “robot teardown”, which allows individuals to learn how to systematically disassemble physical robots in order to learn how to identify vulnerabilities and learn to repair them, thus helping counter planned obsolescence and return some autonomy back to users.

## 3. Factors Affecting Exit in HRI

We define exit in HRI as the emotional or physical separation between a user and a robot. Exit may happen either on a permanent basis (e.g., the robot is no longer available) or on a temporary basis (e.g., a user leaves an interaction with a robot). It may be initiated by a



**Figure 2**

Human-robot-interaction is heavily affected by a dynamic of both engagement and exit forces which affect how a user experiences interaction with a robot. We have identified five domains which influence this dynamic. These include: 1) *situational factors*, such as the cultural context of interaction, the agent’s social role, and the agent’s cultural norms, 2) *physical factors*, such as aspects of the robot or environment which determine whether or not the robot is functioning or successful, 3) *biopsychosocial factors*, which include aspects of the user such as their health, social skills. or mindset, 4) *gatekeepers*, who are individuals which may exert control over when/how/if a user is able to engage or exit an interaction with a robot, and 5) *market factors*, where companies, like gatekeepers, directly affect whether users have access to robots, as well as when those robots may exit due no longer supporting their hardware or software. .

user or external entity, which is an indirect indication of who controls the exit. Furthermore, exit may be intentional or unintentional.

We posit the most important aspect of this process is providing users with agency and autonomy to control both engagement and exit. A part of this is providing tools for mitigating both expected exit (e.g., a desired end to interaction) and unexpected exits outside of the user’s control.

HRI can be considered as a dynamic of engagement and exit forces (see Figure 2). Here, external entities dictate the scope of the user’s engagement or exit with the robot. For example, a company may decide to no longer maintain or support a robot, which forces the user to exit before they are ready to do so. Or, a feature of the environment or context

may force users to continue interacting with a robot even though they wish to exit. A key gap here is that users are rarely able to influence these forces of engagement and exit, nor are they usually aware of them.

There are a series of external entities which influence the engagement / exit dynamic, which we briefly review below.

**Situational Factors:** A variety of situational contexts can have a major impact on an agent's (either human or robot) experience of both engagement and exit. This can include the situational and cultural context of interaction, the agent's social role, and the agent's cultural norms [44, 41, 53, 42]. For example, a user may need to stop interacting with a robot because of needing to attend to another activity or event, or a policy set by an institution may limit interaction for each person to a set period of time. Or, cultural factors around interacting with the robot may cause a user to engage or exit interaction, such some individuals in the middle east being uncomfortable around humanlike robots, or some individuals in Japan feeling an obligation to interact with robots [52, 63].

**Physical Factors:** Physical aspects of the environment or robot can also impact a user's experience of exit. For example, after the 2011 Fukushima nuclear disaster, rescue workers attempted to send robots in to inspect the Daiichi power plant, but the robots quickly melted [56], thus forcing Tokyo Electric Power Company workers to exit prematurely. This also includes a robot not being aware of its physical environment, for example, the Knightscope security robot that did not notice a fountain in Georgetown and plunged into it, thus abruptly ending interaction [17].

**Biopsychosocial Factors:** A whole host of fluctuating interpersonal factors make-up the Biosychosocial factors that are inextricably related to human-robot interactions. These factors encompass the many interpersonal factors that provide internal and external contexts for an individual. These could include a user's physical health status (e.g. to the use of a robotic device for physical support [67]), their social connections and engagements (e.g. connecting with peers through the use of a robot [55]), and also their psychological mindset (e.g. talking to a robot to reduce stress [6]). These factors can also intersect with the technical in increasingly important ways (c.f., the recently proposed biopsychosociotechnical model [11]), which will also affect how users experience exit.

**Gatekeepers:** Gatekeepers often control both engagement and exiting for many populations. Gatekeepers include researchers who may provide access to a robot that would otherwise be inaccessible, parents or teachers who encourage a child to engage with a robot, but may also remove access to the device if deemed appropriate, and even policymakers as in the case of the decision to remove the robotic dog from the NYC police force [74]. Despite these many types of stakeholders who encourage both engagement and disengagement depending upon the circumstances, likely the strongest stakeholder is the Market itself.

**Market Factors:** As a novel and expensive technology, robots are still strongly controlled by their availability and their price. This is true for consumers, researchers, and businesses. Marketing and commercialization of robots necessitates engagement and demand to create profits. In this way, the market is both creating and reinforcing robot demand. Simultaneously, the challenging marketplace has resulted in numerous robots exiting the market due to a lack of funding and profitability. For example, the companies that made Jibo, Kuri, and Pepper all respectively stopped creating and maintaining the robots, leaving users stranded.



**Figure 3**

Two common intervention exits in HRI. An *abandonment exit*, where a robot is permanently removed from use, often without a user’s knowledge or consent. Shibata [60] described several examples of the harm this caused to people with dementia when Paro (left) was removed. In contrast, a *growth exit* is one where a robot intervention is no longer needed. For example, a family using Kaspar (right) for 12 weeks decided their son had outgrown it and no longer needed it.

### 3.1. Examples of Exit in HRI

**3.1.1. Intervention Exit:** Intervention exits are where a robot is taken away at the end of a research study or intervention. This type of exit is most commonly seen in therapeutic and educational applications, where longitudinal deployments are more common. Intervention exits need to be considered carefully given the potential risk of removing active supports that could leave users without other resources, particularly for minoritized populations [21, 34, 28].

**Abandonment Exit:** An abandonment exit is where a robot is permanently removed from a setting, often without a user’s knowledge or consent. For example, Shibata [60] described deploying the Paro robot at nursing homes in Denmark during a longitudinal intervention and rotation program across the country. One resident with dementia had lost their ability to speak, but regained it upon use of the robot. However, after their nursing home’s time with the robot was up, they had to take the robot away, and immediately the resident lost their regained skills. This raises a host of ethical issues, particularly for individuals with cognitive impairments [30, 58, 40, 26].

**Growth Exit:** In contrast, growth exit is where the user has attained valuable skills or practice through the robot and no longer needs the robot’s support. This most often occurs in therapeutic or educational settings. For example, the Kaspar robot [14] was used for 12 weeks at home by a family as a social education tool for their child with autism. A documentary crew followed the family’s use of the robot for multiple weeks, and, by the end of the intervention, the family decided they no longer needed to use the robot, because their child had learned all it had to teach [3].

**3.1.2. Situational Exit:** A situational exit is one in which an environmental, physical, or social situation results in exiting the user from the interaction. This could be strangers or friends interrupting an interaction in a public setting, or a parent overriding a child’s interaction due to a privacy concern. This could also be a situation in which the robot is



**Figure 4**

Situational exits in HRI. A *painful exit* is where a user experiences emotional pain upon exiting interaction. For example, [5] describes a teen who had a difficult time returning to normal life after being immersed in Virtual Reality, and needed to exit (left). A *denied exit* is where a user wants to exit an engagement but is unable, such as the case described by Dr. Emily Ackerman [1], a wheelchair user, being stranded in the street because a delivery robot had stalled out (center). Finally, a *sabotage exit* is where a user either damages a robot or creates barriers in its environment to exit or avoid interaction, such as in the case of Hitchbot’s vandalism (right).

not physically accessible to a user due to its location or position.

**Painful Exit:** A painful exit is one in which the user struggles to exit the technology, resulting in emotional pain. Most parents can imagine this type of exit when witnessing a child’s response to being asked to stop an engaging video game. This type of exit can also be a bit more complex. A recent illustration of a painful exit was described by a teen after using a virtual reality environment for brief periods of time for relaxation [5]. This teen described no longer wanting to use the Virtual Reality (VR) given their discomfort upon exiting. They explained that their real life was so stressful, and the VR environment so peaceful, that returning to real life was incredibly difficult and outweighed the benefit of the experience. Although the technology itself was engaging and pleasurable, the painful exit of the experience resulted in significant discomfort.

**Denied Exit:** Forced engagement is a situation where a user wants to exit an interaction but cannot. For example, Dr. Emily Ackerman, a researcher at the University of Pittsburgh and wheelchair user, was dangerously trapped in the middle of a crosswalk because a Starship delivery robot was stuck on a curb cut, blocking her path [1]. This raises multiple legal questions described by Grimm and Thomasen [20]. Another example is told by Eric Horvitz, where he was stuck in an elevator with a huge delivery robot that nearly injured him when it turned, and he had no way to escape [25].

**Sabotage Exit:** A sabotage exit is where a user either damages a robot or creates barriers in its environment to either exit interaction or avoid it in the first place. In HRI there is evidence of people of all ages sabotaging robots - from children bullying robots in public spaces [27], to middle aged adults vandalizing HitchBot in a park [62], to older adults in an assisted living facility inserting sandwiches into a robot’s LIDAR to prevent it from operating [22].

#### 4. Designing for Exit: Proposed Principles

As evidenced in our examples, potential for harm can result in both unintentional and intentional exits, therefore suggesting that mitigation of these harms is desirable and may be necessary in vulnerable populations. Therefore we propose the following five principles intended for researchers and designers to consider. Unlike design requirements, these are

design principles intended to introduce awareness, discussion, and ideally some mitigation of potential harms resulting from exiting.

**The first principle is *Allow for User-Directed Exit*.** As part of a participatory approach, it is imperative to ensure that users have agency over robot not only participation (whether or not you interact), but also engagements (when and where to have an interaction). This is true across both the commercial sector and HRI research space. Therefore, this principle highlights the importance of ensuring an easy, clear, and feasible exit.

**The second principle is to *Plan for Robot Exit*.** As simple as it may seem, from our own work, and especially what a slew of social robot market exits as shown, it is clear planning for exit is neglected. Therefore this principle invites designers and researchers to have discussions around what exit might look like— both intentional and unintentional—and how those exits may present potential harm to their participants and users.

**The third principle is to *Engage the User in Discussions of Exit*.** This principle stems from a participatory approach and encourages both designers and researchers to explore exit and concerns about exit directly with primary users, particularly DRUs. This principle invites conversation and input from participants about what exit might look like, as well as inviting ideas about possible harms that may not be immediately obvious to designers.

**The fourth principle is to *Expect the Unexpected Robot Exit*.** There are endless examples of exiting that was unintentional and out of the control of everyone involved. This can happen when a company or project loses funding, or when a device becomes damaged and unrepairable. In this case, this principle invites a discussion about how exit is possible and what robot designers and researchers might prepare for, especially if the exit is completely out of their control.

**The fifth principle is *Design for Continuity of Use*.** Similar to the continuity of care model in healthcare, this principle suggests that if a device is providing value (educational, therapeutic) it may be unethical to remove that device - even though a project is complete. For this reason, we invite designers and researchers to consider models that allow for a community of interest to maintain ownership and use of a device that is providing to be beneficial to the community.

## 5. Applying Exit Principles within a Real World HRI project

Each of the four principles of exit can be considered and applied within any robot deployment. How they are applied can vary and be determined based upon other principles and priorities. Here, we give an example from our own work applying the principles to a real world, longitudinal robot deployment.

Project EMAR [39] explores the design and development of a social robot to support teen mental health. Using a participatory approach, we have engaged hundreds of teens in both design and interaction activities to determine both the embodiment and behaviors of the robot. Our social robot is fully customizable, thereby allowing teens to customize and modify the robot to best meet the needs of their community [2]. The first principle of *Allowing for User-Directed Exit* was demonstrated immediately in all of our study designs. Often gathering design and interaction data in situ (in schools or public libraries) we designed interactions and activities that gave full agency to teen participants. We also ensured that teens who chose not to participate in the research (talking to our social robot) could engage in another similar activity (learning how another type of robot was operated)



**Figure 5**

Two teens are engaged in an interaction with the social robot EMAR in a high school classroom.

ideally of equivalent interest. We also designed studies that allowed for teens to exit at any time, for example, each interaction started with “Would you like to share another stress story with me?” to which a participant could select “No” on the robot’s screen if they no longer wished to participate.

The second principle of *Plan for Exit* emerged quickly as one of our first school-based, interaction sessions with teens resulted in a teen expressed a desire to keep the robot at school rather than take it back to the lab with us [7] as we began brief implementations. The third principle of *Engage Users in Discussions of Exit* emerged organically in early design sessions when teens asked, “Can teens talk with the robot if they’re not at school?” Our participants also smartly discussed how gatekeepers (primarily school administrators) may make the robot inaccessible for some students or worse repurpose the robot. Thus blocking a potential interaction, forcing an exit of an interaction (e.g. a student. . .), or removing the teen approved interactions altogether. Through these discussions with teens we developed a requirement in the social robot design that would make it incredible difficult to repurpose the robot. We also developed a plan for “licensure model” whereby school administrators would agree the device remain teen-driven, managed and operated. All of these decisions were direct responses to teens concerns about gatekeeping and unintended exits.

The fourth principle, *Expect the Unexpected*, was discussed frequently throughout our team of faculty and student researchers. Throughout the project we proposed, “What happens when the robot falls down the stairs and breaks?” as a way to continually think forward about our participants’ of potential (and maybe inevitable) technical failures.

The fifth principle of *Design for Continuity of Use* has emerged in a couple of ways throughout the project. Mainly, we quickly realized that removal of the robot, even after brief implementations may disruptive at best, or harmful at worst for our participants. Once in their community, allowing them to continue to use the robot, in any way that is supportive, needs to be considered. Continuity also arises in terms of continuity of use. Will teens be continually allowed access to the social robot to support their mental health? As we consider the many gatekeepers within public spaces such as schools and libraries. Early on in our project we instilled a requirement of teens managing, customizing, and gathering aggregate data from the robot to allow for true agency over the device.



**Figure 6**

When building an exit plan for a longitudinal robot deployment, HRI designers and researchers should: 1) Identify and connect with primary users to understand how they envision engagement and exit, 2) Understand the context of how and where the robot will be used, 3) Identify and make a plan to mitigate factors which will affect sustainability of the robot deployment, 4) Red team for unplanned exits, 5) Empower primary users to create a detailed exit plan, which are then incorporated into the deployment plan.

## 6. How to Build an Exit Plan

Section 5 presents an example of how the five exit principles were considered during the course of a longitudinal robot deployment. For HRI designers and researchers at an earlier stage of their work, it can be helpful to have tools to think through making a plan for exit in advance of a deployment. This is a complex process which involves multiple stakeholders, though can largely be thought of in ways that align with design justice practices, such as participatory action research and emancipatory design [43, 72]. HRI designers and researchers should be working directly with (or ideally integrating) their intended communities during the design process. Ultimately, this approach affords community members the tools they need to have agency over participation, engagement, interactions, and exit.

First, designers and researchers should **identify, connect with, and include direct robot users**. Who will be using this robot? What will they be using the robot for? What might be some barriers to their use (e.g., gatekeepers, situational context, etc). What are the outcomes that DRUs hope for at the end, and what constitutes the end?

Second, it is important to **understand the context of use**. What is the environment the robot will primarily be deployed in? Who has access to the robot? What tasks does the robot have to do vs. what are some hoped-for tasks?

Third, one should **identify and mitigate factors affecting sustainability**. Does the deployment require robot wranglers [64] or wizards [49] in order to be successful? If so, what are some ways designers can minimize the need for wranglers and wizards, and/or transfer those roles to DRUs? How can primary users be empowered to conduct minor repairs on robots when they malfunction?

Fourth, it is critical to **red team potential unplanned exits**. What are some ways the robot might fail? How might the environment (e.g., physical factors) or people (e.g., sabotage) force an exit, and what might be done to mitigate that from an engineering or policy perspective? Here, HRI researchers might consider employing a technique from computers security known as “red teaming”, where one team (the red team) acts as an adversary trying to figure out all the possible unplanned exits of a system, and another team (the blue team) designs potential mitigation strategies.

Finally, HRI researchers and designers should **support primary users to create a detailed exit plan**. Given the red teaming results, what do primary users want to have happen at various stages? Do primary users still want to continue with the deployment given these potential pitfalls? Then, this exit plan should be incorporated into the the deployment plan for the robot.

## 7. Discussion

Participatory research and design encourage us to consider users continually throughout the design and deployment process, in an effort to share decision-making agency and also to mitigate potential harms. Given the numerous domains in which engagement and exiting occur, it becomes essential to take these dynamics into consideration. It is also quite likely that as social robots become more ubiquitous in our daily lives, the types of exiting will only expand, suggesting that designing for exit may be an essential principle to integrate into all stages of both physical and interaction design in HRI.

## REFERENCES

1. E. Ackerman. My fight with a sidewalk robot — national center on disability and journalism. <https://ncdj.org/2019/11/my-fight-with-a-sidewalk-robot/>, Nov 2019. (Accessed on 08/31/2022).
2. P. Alves-Oliveira, M. Bavier, S. Malandkar, R. Eldridge, J. Sayigh, E. A. Björling, and M. Cakmak. Flexi: A robust and flexible social robot embodiment kit. In *Designing Interactive Systems Conference*, pages 1177–1191, 2022.
3. BBC. Six robots and us. <https://www.bbc.co.uk/programmes/b09163d7/episodes/guide>, 2018. (Accessed on 08/24/2022).
4. J. M. Beer, A. D. Fisk, and W. A. Rogers. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2):74, 2014.
5. E. Björling, J. Sonney, S. Rodriguez, N. Carr, H. Zade, and S. Moon. Exploring the effect of a nature-based virtual reality environment on stress in adolescents. *front. Virtual Real.* 3: 831026. doi: 10.3389/frvir, 2022.
6. E. A. Björling, H. Ling, S. Bhatia, and K. Dziubinski. The experience and effect of adolescent to robot stress disclosure: A mixed-methods exploration. In *International conference on social robotics*, pages 604–615. Springer, 2020.
7. E. A. Björling, E. Rose, A. Davidson, R. Ren, and D. Wong. Can we keep him forever? teens’ engagement and desire for emotional connection with a social robot. *International Journal of Social Robotics*, 12(1):65–77, 2020.
8. C. Breazeal. *Designing sociable robots*. MIT press, 2004.
9. C. Breazeal, P. L. Harris, D. DeSteno, J. M. Kory Westlund, L. Dickens, and S. Jeong. Young children treat robots as informants. *Topics in cognitive science*, 8(2):481–491, 2016.
10. J. Bulow. An economic theory of planned obsolescence. *The Quarterly Journal of Economics*, 101(4):729–749, 1986.
11. A. J. Card. The biopsychosociotechnical model: a systems-based framework for human-centered health improvement. *Health Systems*, pages 1–21, 2022.
12. J. Carpenter. *Culture and human-robot interaction in militarized spaces: A war story*. Routledge, 2016.
13. K. Darling. ‘who’s johnny?’ anthropomorphic framing in human-robot interaction, integration, and policy. *Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy (March 23, 2015)*. *ROBOT ETHICS*, 2, 2015.
14. K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. Assif, M. Blow, et al. Kaspar—a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics*, 6(3-4):369–397, 2009.
15. 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 *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 63–72. IEEE, 2020.
16. M. Dziergwa, M. Kaczmarek, P. Kaczmarek, J. Kedzierski, and K. Wadas-Szydłowska. Long-

- term cohabitation with a social robot: a case study of the influence of human attachment patterns. *International Journal of Social Robotics*, 10(1):163–176, 2018.
17. B. Fearnow. Robot suicide? droid security guard drowns in washington dc mall fountain. *International Business Times*.(2017). Available online at: <https://www.ibtimes.co.uk/fukushima-robot-dies-3-hours-after-entering-japans-radioactive-reactor-1496126> (accessed Aug 2022), 2017.
  18. J. Forlizzi and C. DiSalvo. Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 258–265, 2006.
  19. M. A. Goodrich, A. C. Schultz, et al. Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, 2008.
  20. C. Grimm and K. Thomasen. On the practicalities of robots in public spaces. *Proceedings of We Robot*, 2021.
  21. C. Guan, A. Bouzida, R. M. Oncy-Avila, S. Moharana, and L. D. Riek. Taking an (embodied) cue from community health: Designing dementia caregiver support technology to advance health equity. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2021.
  22. M. Hanheide. Social navigation: Problem statement and (some) solutions — dagstuhl seminar 19081, 2019.
  23. W. Hartzog. Unfair and deceptive robots. *Md. L. Rev.*, 74:785, 2014.
  24. F. Heider and M. Simmel. An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259, 1944.
  25. E. Horvitz. Please do not board the elevator with the robot. [http://erichorvitz.com/elevator\\_robot.htm](http://erichorvitz.com/elevator_robot.htm), Nov 1999. (Accessed on 08/31/2022).
  26. L. Hung, C. Liu, E. Woldum, A. Au-Yeung, A. Berndt, C. Wallsworth, N. Horne, M. Gregorio, J. Mann, and H. Chaudhury. The benefits of and barriers to using a social robot paro in care settings: a scoping review. *BMC geriatrics*, 19(1):1–10, 2019.
  27. T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Transactions on robotics*, 23(5):962–971, 2007.
  28. A. D. Kerkhoff, E. Farrand, C. Marquez, A. Cattamanchi, and M. A. Handley. Addressing health disparities through implementation science—a need to integrate an equity lens from the outset. *Implementation Science*, 17(1):1–4, 2022.
  29. A. Kubota, E. I. Peterson, V. Rajendren, H. Kress-Gazit, and L. D. Riek. Jessie: Synthesizing social robot behaviors for personalized neurorehabilitation and beyond. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 121–130, 2020.
  30. A. Kubota, M. Pourebadi, S. Banh, S. Kim, and L. Riek. Somebody that i used to know: The risks of personalizing robots for dementia care. *Proceedings of We Robot*, 2021.
  31. I. Leite, C. Martinho, and A. Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.
  32. Z. Li, L. Terfurth, J. P. Woller, and E. Wiese. Mind the machines: Applying implicit measures of mind perception in social robotics. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pages 236–245, 2022.
  33. J. R. Matisonn. *Stakeholder Models, Sustainability and the Ethics of Planned Obsolescence*. PhD thesis, University of the Witwatersrand, Faculty of Humanities, 2016.
  34. S. Matsumoto, S. Moharana, N. Devanagondi, L. C. Oyama, and L. D. Riek. Iris: A low-cost telemedicine robot to support healthcare safety and equity during a pandemic. In *International Conference on Pervasive Computing Technologies for Healthcare*, pages 113–133. Springer, 2022.
  35. S. Matsumoto and L. D. Riek. Shared control in human robot teaming: Toward context-aware communication. *AAAI*, 2022.

36. S. Matsumoto, A. Washburn, and L. D. Riek. A framework to explore proximate human-robot coordination. *ACM Transactions on Human-Robot Interaction*, 2022.
37. N. Maycroft. *Consumption, planned obsolescence and waste*. PhD thesis, University of Lincoln, 2009.
38. V. Mayoral-Vilches, A. Glera-Picón, U. Ayúcar-Carbajo, S. Rass, M. Pinzger, F. Maggi, and E. Gil-Uriarte. Hacking planned obsolescence in robotics, towards security-oriented robot tear-down. *Electronic Communications of the EASST*, 80, 2021.
39. MeLab. Project emar. <http://depts.washington.edu/melab/projects/project-emar/>, 2022. (Accessed on 08/24/2022).
40. W. Moyle, C. Jones, J. Murfield, L. Thalib, E. Beattie, D. Shum, and B. Draper. Using a therapeutic companion robot for dementia symptoms in long-term care: reflections from a cluster-rct. *Aging & mental health*, 23(3):329–336, 2019.
41. A. Nigam and L. D. Riek. Social context perception for mobile robots. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 3621–3627. IEEE, 2015.
42. A. Nigam and L. D. Riek. Social context perception for mobile robots. in 2015 ieeer/rsj international conference on intelligent robots and systems (iros), 2015.
43. L.-A. Noel. Promoting an emancipatory research paradigm in design education and practice. *Future Focused Thinking - Design Research Society International Conference (DRS)*, 2016.
44. M. F. O’Connor and L. D. Riek. Detecting social context: A method for social event classification using naturalistic multimodal data. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 3, pages 1–7. IEEE, 2015.
45. Y. Pearson. Child-robot interaction: What concerns about privacy and well-being arise when children play with, use, and learn from robots? *American Scientist*, 108(1):16–22, 2020.
46. E. Phillips, X. Zhao, D. Ullman, and B. F. Malle. What is human-like?: Decomposing robots’ human-like appearance using the anthropomorphic robot (abot) database. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 105–113. IEEE, 2018.
47. B. Prusaczyk, T. Swindle, and G. Curran. Defining and conceptualizing outcomes for de-implementation: key distinctions from implementation outcomes. *Implementation Science Communications*, 1(1):1–10, 2020.
48. L. Riek and D. Howard. A code of ethics for the human-robot interaction profession. *Proceedings of we robot*, 2014.
49. L. D. Riek. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1):119–136, 2012.
50. L. D. Riek. The social co-robotics problem space: Six key challenges. *Robotics challenges and vision (RCV2013)*, 2014.
51. L. D. Riek. Healthcare robotics. *Communications of the ACM*, 60(11):68–78, 2017.
52. L. D. Riek, N. Mavridis, S. Antali, N. Darmaki, Z. Ahmed, M. Al-Neyadi, and A. Alketheri. Ibn sina steps out: Exploring arabic attitudes toward humanoid robots. In *Proceedings of the 2nd international symposium on new frontiers in human-robot interaction, AISB, Leicester*, volume 1, 2010.
53. L. D. Riek and P. Robinson. Challenges and opportunities in building socially intelligent machines [social sciences]. *IEEE Signal Processing Magazine*, 28(3):146–149, 2011.
54. L. D. Riek and R. N. Watson. The age of avatar realism. *IEEE Robotics & Automation Magazine*, 17(4):37–42, 2010.
55. R. B. Rosenberg-Kima, Y. Koren, and G. Gordon. Robot-supported collaborative learning (rscl): Social robots as teaching assistants for higher education small group facilitation. *Frontiers in Robotics and AI*, 6:148, 2020.
56. M. Russo. Fukushima: Robot ‘dies’3 hours after entering japan’s radioactive re-

- actor. *International Business Times*.(2015). Available online at: [https://www. ib-times. co. uk/fukushima-robot-dies-3-hours-after-entering-japans-radioactive-reactor-1496126](https://www.ib-times.co.uk/fukushima-robot-dies-3-hours-after-entering-japans-radioactive-reactor-1496126) (accessed Aug 2019), 2015.
57. J. Scholtz. Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, pages 10–pp. IEEE, 2003.
  58. A. Sharkey and N. Sharkey. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and information technology*, 14(1):27–40, 2012.
  59. N. Sharkey and A. Sharkey. The crying shame of robot nannies: an ethical appraisal. *Interaction Studies*, 11(2):161–190, 2010.
  60. T. Shibata. Integration of therapeutic robot, paro, into elderly care in denmark, 2009.
  61. C. L. Sidner and M. Dzikovska. Human-robot interaction: Engagement between humans and robots for hosting activities. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 123–128. IEEE, 2002.
  62. D. H. Smith and F. Zeller. hitchbot: The risks and rewards of a hitchhiking robot. *Suomen Antropologi: Journal of the Finnish Anthropological Society*, 42(3):63–65, 2017.
  63. Y. Sone. *Japanese robot culture*. Springer, 2016.
  64. L. Takayama. Putting human-robot interaction research into design practice. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pages 1–1, 2022.
  65. A. Taylor, D. M. Chan, and L. D. Riek. Robot-centric perception of human groups. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(3):1–21, 2020.
  66. A. Taylor and L. D. Riek. Regroup: A robot-centric group detection and tracking system. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pages 412–421, 2022.
  67. M. Vilela and L. R. Hochberg. Applications of brain-computer interfaces to the control of robotic and prosthetic arms. *Handbook of clinical neurology*, 168:87–99, 2020.
  68. E. W. Wan and R. P. Chen. Anthropomorphism and object attachment. *Current Opinion in Psychology*, 39:88–93, 2021.
  69. A. Waytz, C. K. Morewedge, N. Epley, G. Monteleone, J.-H. Gao, and J. T. Cacioppo. Making sense by making sentient: effectance motivation increases anthropomorphism. *Journal of personality and social psychology*, 99(3):410, 2010.
  70. J. K. Westlund, C. Breazeal, and A. Story. Deception, secrets, children, and robots: What’s acceptable. In *Workshop on The Emerging Policy and Ethics of Human-Robot Interaction, held in conjunction with the 10th ACM/IEEE International Conference on Human-Robot Interaction*, 2015.
  71. B. Woodworth, F. Ferrari, T. E. Zosa, and L. D. Riek. Preference learning in assistive robotics: Observational repeated inverse reinforcement learning. In *Machine learning for healthcare conference*, pages 420–439. PMLR, 2018.
  72. A. R. Wooldridge, T. Nguyen, R. Valdez, E. Montague, M. N. Milner, M. C. Dorneich, and R. Roscoe. Human factors and ergonomics in diversity, inclusion and social justice research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 447–449. SAGE Publications Sage CA: Los Angeles, CA, 2018.
  73. M. Zallio and D. Berry. Design and planned obsolescence. theories and approaches for designing enabling technologies. *The Design Journal*, 20(sup1):S3749–S3761, 2017.
  74. M. Zaveri. Nypd robot dog’s run is cut short after fierce backlash. *The New York Times*. [https://www. nytimes. com/2021/04/28/nyregion/nypd-robot-dog-backlash. html](https://www.nytimes.com/2021/04/28/nyregion/nypd-robot-dog-backlash.html), 2021.