A MULTI-STAKEHOLDER APPROACH TO DESIGNING FOR BEHAVIOR CHANGE IN HCI4D

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Fabian Ndubuisi Okeke
December 2019

A MULTI-STAKEHOLDER APPROACH TO DESIGNING FOR BEHAVIOR CHANGE IN HCI4D

Fabian Ndubuisi Okeke, Ph.D.

Cornell University 2019

Human Computer Interaction for Development (HCI4D) is a rapidly growing community that focuses on understanding and designing technologies for populations that are under-served, under-resourced, and under-represented. Current best practices for designing technologies in HCI4D tend to focus on only the needs of target users, however, technological interventions are often deployed in complex ecosystems that involve multiple groups of stakeholders who interrelate and influence one another. Further, HCI4D researchers often aspire to design technologies that effectively change peoples behavior, however, they rarely engage with foundational theories that are well known in the field of Behavioral Science. As a result, technological interventions in under-served communities do not work as intended and when they do, researchers struggle to achieve sustainability and scalability.

This dissertation addresses these gaps by focusing on how to effectively design technologies that change the behavior of people in complex ecosystems prevalent in under-served communities. Through three case studies conducted in Kenya, New York, and India respectively, I demonstrate how (i) researchers can adopt an approach that engages multiple groups of stakeholders during the design process of a technological intervention and (ii) systematically draw on behavioral science theories to inform technology design. In the first case study, I focus on the design of a health feedback system for beneficiaries in rural Kenya;

the second case study centers on the design of technologies for home health aides who support heart failure patients in New York; and the third case study, presents the design of a low-cost technique for researchers to reduce participant response bias in rural India. By applying the same approach across three highly diverse problem contexts, I show that it is generalizable across problem domains. In doing so, my dissertation makes multiple contributions to HCI4D and builds a stronger bridge between the fields of HCI4D and Behavioral Science.

BIOGRAPHICAL SKETCH

Fabian N. Okeke graduated in 2014 from Fisk University with a Bachelor of Science in Computer Science and Bachelor of Engineering in Computer Engineering from Vanderbilt University. In 2017, he received an M.S. degree in Computer Science from Cornell University. As a doctoral scholar in the Information Science department at Cornell University, his research is in the field of Human-Computer Interaction (HCI) where he combines concepts from Computer Science and Behavioral Sciences to design novel technological tools that change human behaviors. His research primarily focuses on the design, development and evaluation of technological interventions for underserved communities in the United States and in "developing" countries.

To my family for their love, support and sacrifices.

ACKNOWLEDGEMENTS

I am grateful to my special committee (Deborah Estrin, Nicki Dell and Tanzeem Choudhury) who made this moment possible. I am forever indebted to Deborah who took me under her wing and reinforced my confidence when it was extremely challenging at the beginning. I am not only a better scholar but also an improved human with stronger values. I have come to develop the *'DE check'* whenever I am about to take on a task. Also, thank you Tanzeem for welcoming me into your lab space and making me one of your students. Those formative years laid the foundation of this work. Thank you Nicki for coming to Cornell Tech and for becoming my co-adviser—you are my answered prayer in many ways than you can imagine. I always dreamed of working in Africa years after graduating from my doctoral program and because of you I was able to achieve that goal while pursuing my Ph.D. I feel empowered to take on even bigger problems on the continent and I hope to make you proud.

Thank you Cornell for the unique environment that provided the opportunity to interact with several brilliant minds including JP Pollak, Longqi Yang, Eugene Bagdasaryan, Hongyi Wen, Faisal Alquaddoomi, Cheng-Kang Hsieh, Vincent Tseng, Alex Adams, Jean Costa, Saeed Abdullah, Diana Freed, Anthony Poon, Neta Tamir, Aditya Vashistha, Madeline R. Sterling, Benedetta Piantella, and Emily Tseng. Working with each of you made me a better researcher. Special shout out to *‘Chief Scientist’*, Michael Sobolev for becoming both a friend and a critic that strengthened my work. Without you, it would have been impossible to make sense of how to incorporate behavioral science into my research. In addition, thank you to all my collaborators that I worked with while at Cornell particularly Medic Mobile (especially Isaac Holeman and Beatrice Wasunna), Dimagi (especially Mohini Bhavsar), Projecting Health, and Office of Evalua-

tion Services (especially Syon Bhanot). Working with you opened my mind to challenges and opportunities for improving the lives of the world's underserved.

Thank you Diane Levitt for over the years becoming a safe place for me to become grounded whenever mental chaos ensued — and there were lots of them during the program. I feel incredibly fortunate to have had your time, listening ears and warm nature; and I will always be grateful for all these. Thank you Mashfiqui Rabbi for your mentorship throughout the program. Your ability to share both sound advice and spicy food were many times the remedy to Ph.D. hardships. Thank you to all my educational institutions in the last few years (Cornell, Vanderbilt and Fisk universities) for the growth and support particularly for your funding that made this journey possible. Thank you to mentors over the years including Julie Adams at Vanderbilt University for encouraging me to pursue a Ph.D. and guiding me to get into a top program at Cornell. Thank you Sajid Hussain for mentoring me and helping me start my research journey while at Fisk University. Your encouragement over the years have definitely paid off. Thank you to all my friends at Cornell and beyond for your support over the years particularly Vidya Reddy, Yordanos Goshu, Ruke Asagba, Sarah Kayode, Tope Obanla, Tomiwa Lasebikan, and Seye Bankole. Shout out to all of you for patiently letting me vent whenever I felt overwhelmed, eat your free food, and grow together.

Aunty Shade! Thank you. Words cannot fully express how much you have done. More than a decade ago, I thought it was absolutely pointless to daydream about a bachelor's degree in the United States let alone pursue and complete a doctorate. But somehow you saw something bigger—you always have—and convinced my parents and I to take the risk. To me, you're more than a mentor, you are family. Whenever it got too tough I remembered how hard you fought and resolved I couldn’t turn back after all your work. I am eternally grateful for crossing paths with you.

Turning to my champions: thank you to my parents, Sir I.D. Okeke and Lady A.U. Okeke, for constantly being there whether rain or shine. ‘Ijewe iji nwobere’ (slow and steady wins the race). Thanks to my brother, Saviour, for reminding me to always remember the journey so far and for your dedication that made me believe every challenge is surmountable. Thanks to my sisters, Evelyn and Joy, for their persistent belief in me and the ‘keep going Heendu’ affirmations. Thank you Uncle B for your advice, encouragement and support every step of the way. Your tenacity is matchless and kept me strong all this time. Thank you Aunt Chi for your consistent support especially the plenty of jollof rice and soup. In fact, they were the most practical form of help on many occasions. Thanks to my cousins, Ugochi, Udochi and Uchechi who hardly know what a Ph.D. is but still pour out love to me, ‘Uncle BC’. Your voice mails always left me cheerful no matter the occasion.

To my heart, best friend, and die-hard fan, Hina Malik, thank you for coming on this journey with me. I am especially grateful for holding me accountable and motivating me to strive for my best self when it seemed too rough. Here’s to more beautiful journeys. Finally, I am grateful to God for this opportunity. This experience has been a lifetime gift and I would not trade it for anything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This dissertation presents and evaluates methods for increasing the effectiveness of technological interventions in HCI4D. Researchers in Human Computer Interaction for Development (HCI4D), a growing research community within the larger field of Human Computer Interaction (HCI), have focused on understanding and designing technologies for populations around the world that are under-served, under-resourced, and under-represented [54]. Current HCI4D best practices for designing technologies for under-served communities include (see Figure 1.1): (1) develop a relationship with one or more local partners, (2) conduct field research in the geographical location of interest, (3) design a tool tailored to the local contexts by accounting for key factors that impact adoption such as low-literacy, language barrier, lack of internet connectivity, and cost of access, (4) iterate the design of the tool based on feedback from target users until convergence to a final version, (5) deploy the intervention for a specific period of time, (6) evaluate the intervention and discuss future directions. This set of best practices have been adopted over the years in several HCI4D studies [59, 58, 192, 250, 198].

As a concrete example, DeRenzi et al. [57] designed a feedback system for community health workers (CHWs) in rural India. Their work began by establishing relationships with two local partners that focus on improving maternal health in rural India. Then the researchers conducted field research at a local partner site in Uttar Pradesh, India. To develop the feedback system, the researchers shared multiple early designs with CHWs and iterated on them until they arrived at a final version that was implemented on CHWs' basic feature phones. Next, the system was deployed for 12 months with 71 CHWs, and

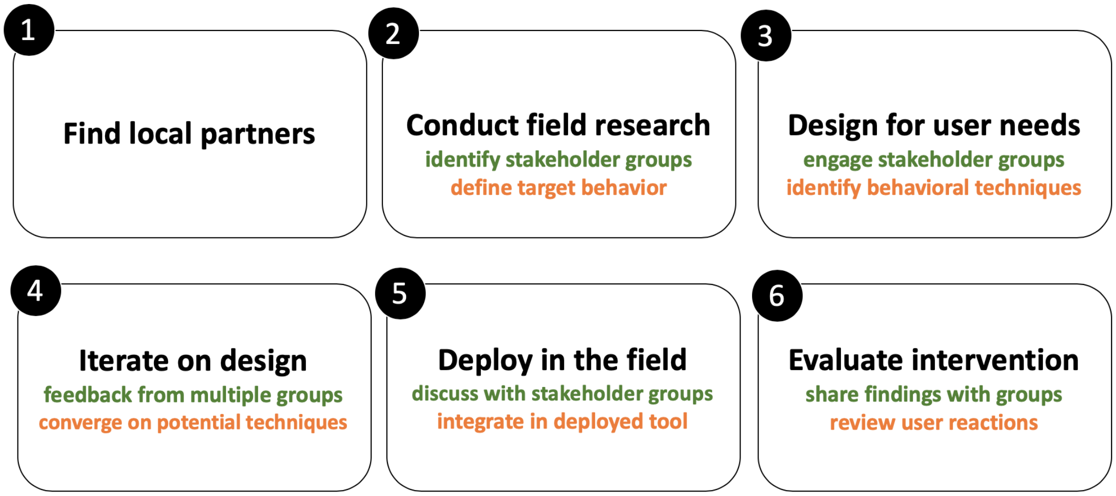


Figure 1.1: The image shows six phases (black text) of current best practices for designing technological interventions in HCI4D. This dissertation extends existing practices from phase two to six by: (i) adopting a multi-stakeholder approach (green text); and (ii) incorporating behavioral theories (orange text) into the design process.

findings revealed that the intervention increased CHW household visits by 20%. Beyond this example, many HCI4D research have utilized similar best practices across a wide variety of research contexts [250, 57, 58, 192, 173].

Adopting a multi-stakeholder approach. Most HCI4D best practices focus solely on the needs of end users (i.e., the people who will actually use a new tool) when designing technologies for under-served communities. However, these technologies are often deployed in complex ecosystems with multiple groups of stakeholders interrelating and influencing one another [89, 54, 138, 219]. Consider for example, researchers create a system that will enable pregnant mothers in rural Kenya to use their basic mobile phones to directly share feedback about the health services they receive from CHWs. Following HCI4D best practices, researchers would engage these mothers to discover their needs (e.g. literacy level, language barrier, internet connectivity) and through iterative testing, improve the design until they arrive at a final version that is implemented and deployed.

Clearly necessary, but we argue not sufficient. What happens when CHWs, who were not part of the design process, learn that there exists a system where users can 'report' CHWs if they do not carry out their jobs as expected? Adding CHWs to the design process to address this issue is critical yet still not enough. Suppose the health center notifies a supervisor that a CHW has been identified for performing poorly — how should supervisors handle this new information considering that CHWs are unpaid and volunteer their time? Beyond supervisors, how do top-level decision makers, such as the director of the health facility and coordinators of local health programs, come into play in the adoption of the new feedback tool?

These thorny questions around engagement of other stakeholder groups beyond end users are not peculiar to the example scenario shared; they also apply to many other research contexts that involve complex ecosystems in under-served communities [89, 18, 138, 244, 236, 219]. For example, to develop technologies for menstrual health education of adolescents in India, Tuli et al. [244] implores designers to *"include an ecology of diverse stakeholders"* by drawing on the perspectives of young adults, parents, teachers, social workers, and health professionals. Although prior work has identified the *why* behind gathering multiple stakeholder-group perspectives, we lack systematic methods for *how* to do this.

My dissertation addresses this gap by showing how to integrate a multi-stakeholder design approach into current HCI4D best practices, particularly during: the development of a relationship with local partners (phase 1), field research (phase 2), when designing a technological artifact that is tailored to local contexts (phase 3), when iterating the design (phase 4), and when evaluating the intervention (phase 6). In each of these phases, my work accounts for the perspectives of different groups of stakeholders in addition to end users.

Incorporating theories of Behavioral Science. Another important concern for researchers designing technologies for under-served populations is how to ensure their designs lead to desired behavior change (e.g. change CHW work patterns [57, 58], engage men in family planning [194], increase women's knowledge about pre-natal care [18]). Although researchers often aspire to design technologies that effectively change behavior, they rarely engage with foundational theories that are from the field of Behavioral Science. Consequently, there is insufficient understanding of the potential ramifications of adopting a behavioral technique.

For example, one study used social comparison charts to motivate CHWs to carry out more household visits [57]. While it encouraged CHWs at the top of the chart, it discouraged those at the bottom consistent with Behavioral Science research that using gaming elements (e.g. comparison charts, leaderboard) can lead to negative effects [241, 93]. A number of studies have applied reminders to change the behavior of target users [58, 192, 198], however, evidence shows that while extrinsic motivators (e.g. a reminder) may work in the short term, in the long run they can reduce people's intrinsic motivation to carry out tasks that they previously enjoyed [53].

Designing behavioral interventions becomes even more complicated when we consider that they are deployed in complex, multi-stakeholder ecosystems as previously described. For instance, an intervention that boosts the number of monthly household visits of CHWs concurrently increases the volume of medications that health centers supply in a month, which in turn multiplies the number of dispensed medications for supervisors to review, and expands the overall health budget for decision makers to plan. Therefore, it is imperative that behavioral interventions for under-served communities are well-thought out and grounded, otherwise they could lead to unintended consequences.

My dissertation extends prior work by showing *how* to incorporate theories and concepts from Behavioral Science throughout the HCI4D design process. In particular, researchers can adopt a behavioral lens: during field research (phase 2), when designing a technological artifact that is tailored to local contexts (phase 3), during design iteration (phase 4), and when evaluating the intervention (phase 6). In each of these phases, I systematically draw on Behavioral Science theories to inform the design process.

In the following chapters, I illustrate my general research approach that (1) takes into account the perspectives of multiple stakeholder groups in HCI4D contexts, and (2) incorporates Behavioral Science theories into technology design for under-served communities. I apply this approach to three diverse case studies conducted across different countries, target users, partner organizations, and research domains, each of which I describe briefly below.

Designing a feedback system for care recipients in Kenya (Chapter 3). In the first case study, I designed a feedback system for care recipients in rural Kenya. Consistent with HCI4D best practices, our research team began by establishing relationships with multiple local partners in Kenya including the Ministry of Health (MoH), Medic Mobile and Living Goods, two NGOs, and the county leadership in our research site. Then we conducted a qualitative research study that revealed current feedback practices and barriers faced by care recipients, CHWs, and their supervisors [181].

Building on insights from our qualitative research, we explored the design and deployment of a feedback system based on USSD (Unstructured Supplementary Service Data) that allows anyone who possesses a basic mobile phone to provide feedback regarding the health services and quality of care they received from a CHW or during a hospital visit [185]. Although care recipients were our target users, we adopted a multi-stakeholder approach by also engaging CHWs and supervisors as key stakeholder groups throughout the design of our intervention. For instance, involving CHWs in the design process enabled them to perceive the tool as an avenue for supporting accountability and not for punitive measures.

In parallel, we integrated a behavioral lens into our design process by first performing a review of behavioral techniques used in existing feedback systems around the globe. Our review of over 1100 documents (including peer-reviewed research and organizational reports), revealed three dominant techniques used in digital feedback systems: *positive reinforcement* [222], *framing effect* [245], and *social proof* [217]. However, we applied only *positive reinforcement* because it did not require building up on a user's historical data like the other two concepts.

Further, we built on the theory of *choice architecture* to inform the placement of contents in our system because there is evidence that the order of presentation of items impacts people's decision making [11, 117]. For example, our tool showed users simple questions first (e.g. your gender) before asking them to share their feedback. After implementing the system, we deployed it with 42 participants for seven weeks in Siaya, Kenya. Findings from our deployment showed that USSD is a promising approach for enabling care recipients to submit feedback in a way that balances privacy, equity, and sustainability.

Designing to improve equity for home health aides (Chapter 4). The second case study designed a technological intervention for home health aides (HHAs) in New York City (NYC) who support patients with heart failure, a life-threatening disease with unpredictable symptoms. Similar to the first case study, we established relationships with relevant local partners: three agencies that employ

HHAs; and two hospitals where patients are treated for heart failure. Since little is known about the current impact of technology on HHAs' work practices, we conducted a qualitative research study in NYC to map out the actors in the heart failure ecosystem and identified eight groups of stakeholders: patients, family care givers, HHAs, physicians, nurses, social workers, care coordinators and agency leaders [184].

Our findings revealed that HHAs are marginalized and undervalued in the heart failure ecosystem and we ideated on ways in which technology could alleviate their challenges. Building on this, we constructed a design provocation: a non-functional prototype of a tablet application that addressed HHAs' needs and tested it with different stakeholder groups to reveal the dilemma of equity inherent in this complex ecosystem [243].

In addition, we reviewed Behavioral Science literature for techniques that directly addressed decision making, a key challenge for HHAs. In our review we considered *nudge theory* [240], *knowledge learning theory* [86] and *simple heuristics* [46]. Although the last two concepts were relevant, they were more appropriate after a heart failure curriculum had been designed for HHAs. Building on *nudge theory*, we added in our provocation, immediate feedback responses delivered in a background color that was either *green*, *yellow* or *red* alert signals, as a way to guide HHAs in their work. For example, suppose an aide is reporting their daily tasks in the tablet and they indicate that a patient fell but still looked fine, the app would display a message in *red* background prompting: *"contact your supervising nurse immediately"*. Designing the provocation in this way provided insights on nuances to account for when designing future interventions in this ecosystem.

Designing to reduce participant response bias in HCI4D (Chapter 5). Finally, the third case study focused on the design of a practical, low-cost intervention to mitigate participant response bias. Compared to the other two case studies that each began with qualitative research to understand an ecosystem, this study took place in a context that was well understood with multiple research already conducted in the past. It is well-known in the HCI4D community that in the evaluations of new designs or technological artifacts, researchers find it challenging to elicit critical feedback from participants in under-served communities due to participant response bias [186], which stems from large social and cultural differences between researchers and participants [108, 3, 101, 146, 55].

To address this problem, we designed a practical, low-cost intervention and tested it in two controlled experiments (online and in the field) that successfully demonstrated that our intervention mitigated participant response bias [255]. We began our research by reviewing Behavioral Science literature for ways to persuade users to share truthful feedback. In addition, we consulted two behavioral experts for suggestions on addressing response bias. To narrow down our list of potential behavioral methods, we eliminated techniques that were challenging to implement and difficult to scale with technology, and arrived at *priming* [226] and *social proof* [217] as promising techniques. Next, we tested each of these concepts in a controlled online experiment, which revealed that *social proof* effectively elicited critical feedback from users. Building on the findings of the online experiment, we conducted a field experiment with 63 women in rural India, which yielded similar results demonstrating that social proof can mitigate participant response bias in under-served communities.

In addition, we adopted a multi-stakeholder approach by partnering with Projecting Health, a local organization focused on community-driven interventions to improve maternal health in rural India. The staff at the program produce videos on health education but struggle to receive critical feedback from their users, marginalized women, on how to improve these videos. Therefore, they were interested in the impact of our intervention. As a result, we checked back with them during our evaluation phase and they shared that our intervention encouraged the women in their program to provide relevant and actionable feedback that led to a three-day video production training for a number of staff.

Systematically combining a multi-stakeholder view with a Behavioral Science lens to three diverse problem contexts, this dissertation demonstrated generalizability across problem domains and made the following contributions:

- The first case study designed a universally accessible feedback system that applied USSD for the first time to the domain of community health. This work addressed an important gap in the community health literature, which has thus far focused on collecting feedback from CHWs, supervisors, and decision makers but not care recipients [181, 185].
- The second case study analyzed how technologies could be designed in ways that improve equity for formal, paid home health aides who care for heart failure patients in the United States, a "developed" country. This work expands HCI4D literature on community health that has so far focused on low- and middle-income countries [184].
- Third case study developed a practical and generalizable technique for HCI researchers to influence participant response bias and encourage participants to provide critical feedback when evaluating an artifact [255].
- In aggregate, these studies demonstrated how to account for the perspectives of multiple groups of stakeholders, beyond only end users and doing so illustrated how to design technologies more effectively in complex

ecosystems for under-served communities.

- Finally, this work sheds light on how to systematically incorporate theories from Behavioral Science to inform the design of technological interventions across diverse contexts. In doing so, this work helped to build a stronger bridge between the fields of HCI and Behavioral Science.

CHAPTER 2

RELATED WORK

This dissertation sits at the intersection of two research areas, HCI4D and Behavioral Science, with a focus on effectively designing technologies in a way that engages multiple stakeholder groups and promotes behavior change in under-served communities. To this end, this chapter discusses relevant literature from both research areas, specifically on the engagement of end users during the design of technologies and on incorporating Behavioral Science theories into the technology design process. While this chapter provides an overarching related work for the entire dissertation, chapters 3, 4, and 5 that describe three case studies also discuss relevant literature specific to the domain of each case study.

2.1 The Context and Settings of HCI4D

The term 'HCI for Development' or 'HCI4D' was coined in 2007 by Chetty et al. [39]. HCI4D is concerned with adapting traditional HCI practices to design technologies for populations that are under-served, under-resourced, and under-represented around the world. Seminal work by Brewer et al. [28] was one of the first studies to lay the foundation of HCI4D. Their paper highlighted that *"even the basic components of computing interfaces encounter problems in developing regions"* and called attention to the importance of engaging with HCI practices when designing technologies for people in developing regions. However, the authors emphasized that technology cannot be seen as a universal remedy for the complex problems that under-served communities face; rather, researchers and practitioners deploying technological tools should follow a well-thought-out approach that accounts for socio-technical issues impacting a target population.

Embodying this sentiment, HCI4D researchers do not take a technology-first approach but instead seek to first gain a deeper understanding of the contexts of users in under-served regions and the nuances involved in using technologies to address the societal issues that impact these users. For example, Kumar and Anderson [137] studied the mobile phone practices of women in rural India engaged in a public health initiative that created health-related short films to address maternal and infant mortality. Their research revealed that despite operating within strict patriarchal norms with limited freedom, these women were able to mobilize help from children and youths to use their mobile devices in accomplishing various tasks. Building on this insight, the authors identified effective ways of disseminating health information through technology and stressed the importance of *"looking not only at users but the context they operate within — including the technological and human elements that influence their social and cultural activities" [137].*

Since the formation of HCI4D, a number of research papers have reviewed the growth of this research community [3, 35, 54, 101]. Ho et al. [101] reviewed 65 articles to unfold the progression of HCI4D-related work from as early as 1982 to 2009. Their review discussed the fast-growing nature of the community, how the cultural characteristics of local communities impact technology design, the participation of community members during the design of a technology, the adoption of user-centered design as a primary design approach, and key issues affecting the growth of the discipline. In particular, they emphasized the tension in balancing research activities (e.g. publishing multiple papers to advance one's career) with development practices (e.g. scaling a promising technology even if it might not contribute to the originally-stated research goals).

At about the same period, other researchers reflected on the HCI4D journey so far. Anokwa et al. [3] documented the experiences of nine HCI4D researchers who collectively conducted research across 13 low- and middle-income countries in three continents and diverse research domains such as agriculture, health, and education. Their work highlighted the challenge of managing the expectations of research partners and echoed similar tensions in balancing research and development goals at the same time. Following this, Dell and Kumar [54] conducted a review that involved a two-pronged approach: a literature review coupled with an 'insider' perspective based on interviews with HCI4D experts. They reviewed 259 HCI4D papers between 2009 and 2014 to get an overview of the growth of HCI4D since its last review. This revealed that about 90% of HCI4D research took place in Asia and Africa with approximately half of all research happening in Asia. It also shows the rapid adoption of mobile phones in research studies, and education, access, and health as the most prevalent domains studied.

In the second part of their work, they interviewed 11 experienced HCI4D researchers to capture *'insider'* sentiments about the trajectory of the community. Findings from their analysis revealed that HCI4D is perceived as an HCI subfield interested in making *"technology more accessible to people who have so far been excluded"* and its reach can be further broadened by expanding the 'D' in HCI4D to encompass *"development outside of foreign lands far away"* including under-served communities in the United States and in Europe. This dissertation aligns with this vision as it studied under-served communities in not only Kenya and India, as developing nations, but also in the United States. Further, their work emphasized that as the HCI4D community continues to grow, there is a need for stronger partnerships between researchers, practitioners, policy makers and governments. This doctoral research embodies this recommendation as all three case studies presented (in chapters 3, 4, 5) were conducted through strong partnerships between researchers, practitioners at NGOs, decision makers at local government counties, hospitals, and health agencies. I now discuss literature relevant to a key approach in this research: engaging multiple stakeholder groups during the design of technological interventions in HCI4D.

2.2 Engaging Multiple Stakeholder Groups in HCI4D

An important best practice in HCI4D is to tailor the design of a technological artifact to the needs of end users and account for factors that could impact adoption such as low literacy, language barrier, and internet connectivity. It is also important to iterate on the design of the tool based on feedback from target users until a final, usable version is achieved. As an example, Medhi-Thies et al. [166] undertook a multistage design process when developing a social network mobile application for low-literate farmers in rural India. Their app accounted for a number of design constraints such as adopting audio interaction rather than textual content because farmers could not read; connecting everyone on the social network to everyone else to mirror village life; and using static images in place of videos due to low bandwidth internet connectivity. Based on feedback from farmers, they iterated on their design twice before finally deploying it.

In another example that focused on the design of a feedback system for CHWs, researchers tested diverse ways for CHWs to visualize their personal performance [57]. Each design phase began with an introduction (on paper) to bar graphs, line graphs, and pie charts to ensure proper understanding of visualizations. Then CHWs saw an early mockup of a barchart that showed total number of visits per week for five weeks. The choice of colors for charts also played an important role in cultural interpretations (e.g. green was perceived to be a beauty queen and yellow as a dirty fellow). After iterating on over 20 designs, the researchers converged on a final set of visualizations that were deployed in the system. Consistent with current HCI4D best practices, these two examples focused on the needs of the end users – farmers and CHWs respectively.

However, prior work has shown that ecosystems in under-served communities are complex environments that consist of multiple groups of stakeholders who influence one another [89, 54, 138, 219]. As a result, the impact of a technological intervention is not limited to only the specific end users designed for; rather, other stakeholder groups are affected. For example, farmers in the social network system previously described could only access their accounts when field extension officers with mobile phones visited them and served as technology mediators [166]. The mediators in turn had translators who called in to encourage them to use the system. Although the system was designed as a social network, farmers struggled to perceive it as one partly due to the dependence on their mediators who did not perceive the system as a social network despite receiving extensive training prior to deployment. Therefore, the influence of mediators, another stakeholder in the ecosystem, impacted how farmers regarded the application.

Comparably, in the previously described design of a personalized system for CHWs [57], the perspectives of supervisory staff were not accounted for even though they are directly impacted by the performance of CHWs. For instance, an increase in CHW visits implies a rise in the number of activities supervisors ought to monitor as well as a jump in the number of supportive household visits that supervisors make. Batool et al. [18] revealed that a mobile phone intervention for low-income mothers in Pakistan to increase their health education was impacted by complicated family dynamics involving mothers-in-law. For instance, mothers-in-law prevented mothers from visiting hospitals unless they had complications. But the perspectives of these stakeholder groups was not factored in because the intervention only focused on mothers.

As a final example, Poon et al. [198] designed an SMS-based intervention that assisted high school students in Cameroon in preparing for their upcoming exams through daily quiz questions received via WhatsApp. However, the researchers noted that when a parent sees their child using WhatsApp, it is not immediately evident that the child is studying; instead they appear distracted. As a result, the authors recommended that future educational interventions should include parents' perspectives because *"parents' ability to place control on students' digital activities can reduce participation"* and may have resulted in lower answer rates observed in the study.

These studies point to the growing need to look beyond the needs of only end users and account for the perspectives of other stakeholder groups in these complex ecosystems. Although recent research demonstrates a growing interest in accounting for multiple groups of stakeholders when designing technologies for under-served communities, to date these efforts have not been performed in a systematic way. For example, to develop technologies for the menstrual health education of adolescents in India, Tuli et al. [244] highlighted the need for designers to *"include an ecology of diverse stakeholders"* by drawing on the perspectives of young adults, parents, teachers, social workers, and health professionals. The authors emphasize that while menstrual health education primarily targets women, it is critical to gather multiple perspectives because communication around this subject takes place in different learning contexts including schools, homes, and workshops.

In a qualitative inquiry into clinics introduced by the government of Delhi (India) to improve healthcare access, Ismail et al. [111] studied who the intended stakeholders of clinics were and how they engaged with the health facilities. Their findings revealed that clinics were unable to effectively address the needs of patients because doctors made assumptions about patients' lifestyles and did not pay attention to the perspectives of CHWs. As a result, the authors recommend that designers of healthcare interventions bridge the disconnected perspectives of *all* stakeholder groups including healthcare providers, CHWs, and patient groups. Although identifying stakeholders in an ecosystem is an important step, it is not the only step. Too often, stakeholder groups are identified in the early scoping of a research study but researchers do not follow through and engage with them during the design of an intervention. For instance, one study focused on the design of a phone-based intervention for adults in Peru to earn high school degrees through distance learning [30]. During the initial phase of the project, school officials were engaged and approved of the system before and during its implementation but they were not involved in the design details. The outcome was that the project was successfully set up and became operational but teachers and principals used their phones to communicate with each other instead of for the intended distance learning.

Kumar et al. [138] discussed lessons learned from the deployment of three mobile health projects to assist CHWs in India, Kenya and Zimbabwe. Their work posited that a design must go through three stages (in-lab, in-field, and scaling) to extend beyond deployment. In the first stage, a proof-of-concept is developed and a partner organization is identified. In the second phase, the technology is tested with the target user, and in the third phase, scaling is planned. The authors acknowledged that while the first two stages were successfully accomplished, each project struggled with the third stage. As a result, they emphasized that designers *"must factor various stakeholder concerns in all three stages—in-lab, in-field,* *and designing for scale."*

Although prior work has identified the relevance of gathering multiple stakeholder-group perspectives, existing efforts have mostly occurred when identifying actors in an ecosystem or as an after-thought after an intervention has already been deployed. To date, there is no clear roadmap on how to systematically engage multiple groups of stakeholders from the inception of a project to its deployment and scaling. This dissertation fills this gap by illustrating how researchers can adopt a multi-stakeholder approach by consistently drawing on (i) the perspectives of specific end users and (ii) the viewpoints of other stakeholder groups throughout the design and deployment of technological artifacts.

For example, in one of the case studies presented on developing a health feedback system for care recipients in rural Kenya (chapter 3), our research team began by establishing relationships with two NGOs, county leadership, decision makers at hospitals, CHWs, supervisors, and care recipients. Then we discussed scaling efforts and the sustainability of potential feedback mechanisms at the inception of the project and long before deployment. During our iterative design phase, we intentionally engaged with these diverse stakeholders and improved our design based on their feedback. Finally, after deployment and during evaluation, we again engaged with these different stakeholder groups, sharing our findings with them. By adopting the same approach across three case studies (chapters 3, 4, 5), this doctoral research provides a practical way forward for HCI4D researchers to formally combine the perspectives of end users and other stakeholders in a specific ecosystem. In doing so, this dissertation expands the current best practices for designing technologies in HCI4D.

2.3 Designing Technologies for Behavior Change

Researchers in HCI4D frequently aspire to design technological interventions that change people's behavior. However, they rarely engage with foundational theories in the field of Behavioral Science to inform their intervention design. Outside of HCI, a growing body of Behavioral Science research has explored the adoption of diverse behavioral interventions in developing contexts, often in the form of evaluation studies or randomized controlled trials. These studies, deployed as large-scale interventions, usually cover hundreds or thousands of users across diverse domains including health [13, 224], finance [10, 31], and education [12, 78].

For example, one study assessed the effectiveness of a commitment savings bank account to encourage more financial savings in the Philippines [10]. Clients of a rural bank (n=1777) were randomly assigned to receive a new savings account or not receive one at all. For clients who received new savings accounts, each account came with restricted access to deposits and did not compensate the client for this restriction. After 12 months of use, the average bank account savings for the new account group increased by 411 pesos relative to the control group and 81% compared to pre-intervention savings levels, suggesting that pre-commitment is a promising behavioral technique for saving money.

There is evidence that, beyond infrastructural problems that the poor encounter, they are also prone to cognitive biases that impact their behavior. In their work on how to design behavioral policy programs, Datta and Mullainathan [50] explained that although people in impoverished communities *"live under a very different set of economic constraints than their counterparts in richer countries, they face very similar cognitive and psychological constraints as their richer counterparts—but with fewer institutional aids to help them overcome the consequences of these common* *constraints."* As a result, the researchers emphasized the need to additionally study the behaviors of the world's poor through well-established theories on human biases as this could provide insights on how to design effective programs.

For instance, to combat the spread of malaria in rural Zambia, mosquito nets were provided to inhabitants of a rural community [82]. Instead of using the anti-malaria nets as intended by health experts, many villagers combined the nets provided to their family members and sewed them into gigantic sieves used for fishing. It is not that the beneficiaries of this public health initiative were unaware of the benefits of the anti-malaria campaign, rather it is due to *present bias* [179]—the immediate pressing needs of these individuals to address hunger does not align with their future need, to prevent death by malaria. One villager embodies this sentiment: *"I know it's not right but without these nets, we wouldn't eat."* Failing to address this sort of behavioral barrier results in wasted resources and does not tackle the original problem that motivated the initiative.

To combat the human cognitive biases of impoverished communities, researchers have explored diverse behavioral techniques. For example, Duflo et al. [71] identified that a key reason why farmers in rural Kenya experienced low harvests was because they suffered from *present bias* [179] i.e. they procrastinated on when they would purchase fertilizers, what type of fertilizer they needed, and how much they planned to purchase. To address this behavioral problem, the researchers designed a program that offered free delivery of fertilizers to farmers early in the season. By bringing the fertilizers to farmers, the researchers hypothesized that it would reduce the tendency of *"planning to do it later"*, which often resulted in no action. Results from the intervention revealed an increased fertilizer use by up to 70%, suggesting that a time-limited small discount program may be an effective policy to encourage fertilizer use.

Behavioral researchers have also begun to use mobile phone technologies in their intervention designs due to the increasing adoption of mobile devices across the globe. For instance, a three-arm randomized controlled trial in Kenya tested the impact of an SMS intervention on facility delivery, exclusive breastfeeding, and contraceptive use [247]. Three hundred women participating in antenatal care were randomized into one of three groups and monitored for 24 weeks postpartum. In the control arm, women received routine clinic-based counseling and care. In the second arm, participants received one-way SMS messages every week that educated them about maternal health. In the third arm, users had access to a two-way SMS where they could interact with nurses. Results from the study showed that while the impact of facility delivery did not differ by study arm, the SMS intervention increased exclusive breastfeeding practices and early contraceptive use; and when the SMS was interactive, the effects on breastfeeding were sustained.

Although relevant, prior behavioral research focused on low-income communities often adopts a top-down, one-size-fits-all intervention approach that targets users across multiple communities while ignoring their local contexts. In addition, only a small number of these studies adopt technology as a potential avenue to scale interventions and, when technology is used, the intervention design does not follow a rigorous, bottom-up approach that iterates on the design and tailors the tool to people's needs. By contrast, HCI4D researchers that aim to design behavior change interventions for under-served communities situate their designs to the context and specific needs of end users.

For example, one study examined the use of persuasive messages to change the habits of women in rural India [202]. During preliminary qualitative research, the authors learned that women in the villages ignored health advice from CHWs and this in turn impacted the motivation of CHWs. Building on persuasion theories, the researchers created short persuasive videos that highlighted the dangers of specific health diseases and how to take preventive measures. CHWs were also encouraged to partake in video creation. Then they shared these videos on their mobile phones during household visits with clients. Findings revealed that the videos improved CHWs' engagements with their clients, increased CHW knowledge of health practices, and boosted their intrinsic motivation.

However, technological interventions in HCI4D need to be carefully designed as they could lead to negative consequences due to the complexity of underserved ecosystems. For example, Lazem and Aly Jad [145] explored the extent to which digital technologies could motivate primary school students in Egypt to learn a challenging topic in Mathematics. To this end, they designed an educational game that supported students to form team groups that worked together to memorize multiplication facts. Findings from the research show that students improved in how they learned and achieved this with minimal supervision from their teachers. However, the intervention created competition tensions between students and enabled cheating behaviors.

A consistent theme in such studies, and broadly in HCI4D research, is that the complexities of the ecosystem of interest could impact scaling. This is in part due to focusing intervention designs on only end users without accounting for the potential effect on other stakeholder groups. It may also be because intervention designs do not happen in a systematic way that accounts for the potential consequences of adopting a behavioral technique. As a result, interventions often do not work as intended or they work but bring about undesired outcomes.

To shed light on unintended consequences, it is essential to apply a behavioral lens to the design of a technological intervention. After all, the intervention is targeting the change of a specific behavior. For instance, an intervention that serves as an extrinsic motivator (e.g. educational games, SMS reminders) could lead to short term gains, but it in the long run they could crowd out people's intrinsic motivations to carry out activities they enjoyed before the intervention (e.g. games resulting in a low desire to learn in classrooms, reminders leading to low motivation for CHWs to carry out household visits). This phenomenon has been well documented in the behavioral science literature as the *crowding out effect* [53] but is rarely considered in HCI4D research.

As another example, several HCI4D interventions send messages that try to nudge users to take specific actions. However, there is little reference in HCI4D literature to *nudge theory* [240] and how digital content could be designed to guide users towards a specific direction. This has real-life implications. For instance, an educational intervention that improves student learning could discourage teachers because of increased workload. Similarly, an intervention that increases CHW performance could receive push back from supervisors when it is scaled because of increased supervisory activities. However, building on *framing effects* [245], as part of *nudge theory*, the additional workload could be respectively framed as *"a teacher has saved a young child from early marriage"* or *"a supervisor has assisted in mitigating the death of a newborn baby."* Drawing on behavioral techniques in this way by accounting for both target users and other stakeholder groups could reduce unintended outcomes in technological interventions.

This dissertation bridges this gap in HCI4D by presenting a systematic approach for incorporating Behavioral Science theories into the design of technological interventions for complex ecosystems with multiple stakeholder groups. This approach includes assessing a research problem to understand the cognitive biases involved in the community of interest, drawing on behavioral science literature to highlight potential ways to address the issue, converging to specific behavioral concepts that can be implemented via technology, and implementing and evaluating the technique.

In the next three chapters (3, 4, 5), I present three case studies that each describe in detail how I used my approach to engage with multiple groups of stakeholders and incorporated Behavioral Science theories into the design of new technology interventions in HCI4D settings. The first case study focuses on the design of a health service feedback system for care recipients in rural Kenya. The second case study centers on designing technologies for home health aides supporting heart failure patients in New York. The final case study, which took place in India, presents the design of a low-cost technique for mitigating participant response in under-served communities. Chapter 6 then synthesizes these case studies into broader takeaways, highlighting lessons learned and concluding with a path forward for future HCI4D scholars.

CHAPTER 3

CASE STUDY 1: A HEALTH SERVICE FEEDBACK SYSTEM FOR CARE RECIPIENTS IN RURAL KENYA

3.1 Introduction

This chapter presents the design, deployment and evaluation of a universal system that enables care recipients to use any mobile phone to share feedback about the health services they receive in rural Kenya. Community health programs in low-resource environments (like rural Kenya) provide essential health services to vulnerable populations. Well-functioning community health programs receive input from and pay attention to the needs of multiple groups of stakeholders, including community health workers (CHWs), supervisors, government ministries, NGOs, and, of course, the communities of people who receive care. Unfortunately, prior work has shown that the quality of the community health services that are delivered is not always on par with what care recipients should receive. For example, a study at a referral hospital in Tanzania found that 70% of women experienced disrespect or abuse during labor and delivery [211]. To try and address such problems, health programs are interested in collecting data that might promote accountability, transparency, and equity [102]. A first step in this direction is to design new tools that encourage individual care recipients within target communities to voice their opinions about the health services they receive.

The HCI4D community has a growing interest in understanding the complex socio-technical systems that impact communities in the Global South [266, 250, 57] and the design of technologies that engage multiple stakeholders in these communities [266, 177, 258, 223]. In the community health literature specifically, prior research has looked at how to gather feedback from CHWs, supervisors, and high-level decision makers [173, 57, 58], but there remains a need for research that examines how to collect feedback from *care recipients*. Recent research reinforces this need by discussing the potential benefits and opportunities for new feedback systems that target care recipients [181].

This work begins to fill the gap in the community health literature by designing a new system that enables community members in rural Kenya who possess only a basic mobile phone to submit feedback about care received during a CHW or hospital visit. The system, which is available in three languages (English, Kiswahili, and Dholuo), was designed through an iterative, stakeholder-engaged approach that included the opinions of care recipients, CHWs, and high-level decision makers.

We implemented the system using USSD, a universal communication channel available on any mobile phone (e.g. users dial *144# to check their airtime balance on the Kenyan network Safaricom). Although USSD is universally available and not a new technology, it has been surprisingly overlooked as a potential channel for engaging underserved communities in the Global South [193]. To date, USSD has primarily been utilized only for carrier service requests and mobile money transactions [197]. Our research expands the limited literature on USSD to a new context and, to the best of our knowledge, is the first to apply USSD to the domain of community health.

During our design process, we engaged with Behavioral Science literature to gain insights on how to design the system to be more persuasive and easy-to-use. After designing the system, we deployed it for seven weeks in Siaya, a rural county in western Kenya, collecting quantitative system usage data as well as qualitative data from six focus groups with 42 participants. Our findings show that 168 care recipients engaged with the system during our deployment, submitting 495 reports. Most of the feedback received was positive and submitted in Dholuo, the local language. We show how sending SMS reminders to care recipients triggered engagement, and highlight a range of socio-technical factors that impact our system, including the importance of preserving user privacy and promoting equity by making the system free to use. Finally, we discuss key themes to address as we move forward: the feasibility of USSD as a mechanism for collecting feedback from communities in low-resource settings, the actionability of such feedback, how we might scale the system, and challenges impacting sustainability. To summarize, our contributions to the HCI4D community are:

- We describe the design and deployment of a mobile phone-based system that collects feedback from *care recipients* in community health programs in the Global South. In doing so, we address an important gap in the community health literature, which has thus far focused on collecting feedback from CHWs, supervisors, and decision makers.
- We expand prior literature on USSD [193] by being the first to apply USSD to the domain of community health. Our findings suggest that USSD is a feasible mechanism for collecting health-related feedback from rural communities in the Global South.
- We discuss a diverse range of socio-technical factors that impacted the design and deployment of our feedback system, revealing important tensions that arose as we strove to create a system that balances privacy, equity, and sustainability. These insights could inform how researchers and practitioners adopt and use USSD in low-resource contexts.

3.2 Related Work

3.2.1 Feedback in Resource-Rich Contexts

The impact of feedback has been extensively studied in resource-rich contexts, particularly, in shared economy such as eBay, Airbnb and Uber [129, 167, 84, 157]. These online platforms allow users to publicly share their opinions about past experiences using ratings and written reviews that create a reputation environment for building trust between participants and service providers [237]. This provides benefits for both the users and service providers: users can build on feedback provided by others to form expectations of trustworthiness when choosing a service; while high reputation providers are considered more trustworthy and can accrue increased financial benefits compared to low reputation sellers [167, 84]. Beyond the sharing economy, recent research has explored how feedback systems can engage people who use the services of care organizations in resource-rich settings [70, 69]. While undoubtedly relevant, two major contextual differences limit the confidence with which we might generalize these findings and design approaches to our current project. The first has to do with institutional arrangements and routine practices that differ between e.g., online shopping and delivering public health services. The second has to do with broader socio-cultural dynamics that may shape and constrain the possibilities for feedback in its various forms.

3.2.2 Feedback in LMIC Services

For decades, citizen movements in LMIC have advocated for increased transparency, accountability, participation, good governance and effectiveness in the major public and private institutions that serve them [131]. More recently, ICTD researchers have begun to explore how digital technologies are playing a role in these efforts. For example, grievance redressal systems have been deployed as part of government accountability and transparency initiatives used to collect citizen responses through telecenters [201], web portals [171, 204] and IVR systems [36, 225]. Specifically concerning health institutions in LMIC, a recent review identified a range of interventions and organized them into four categories related to: 1) gathering citizen feedback; 2) visualizing governance problems; 3) mobilizing for change; and 4) addressing fraud or corruption through automation and auditing [102]. Some of these projects have reached a large scale; for example, UNICEF's community empowerment platform U-Report boasts over five million users worldwide [246]. The maternal health messaging service MomConnect in South Africa has registered over half a million women and had gathered over 4,000 compliments and 690 complaints as of 2016 [17]. A repeatedly recognized shortcoming in this work is a tendency to focus on gathering information to promote transparency, while neglecting or experiencing significant challenges in the work of *using feedback* to improve the quality or equity of services [102, 17]. For example, the Bophelo Haesoa pilot study in Lesotho undertook an extensive, participatory design process that explored new ways for nurses to use apps and organize skits to gather community feedback, but the paper pays relatively little attention to the use of feedback in improving the health system [173]. This is rather ironic, because functioning feedback systems that use data to improve health worker performance are now widespread, and some of them are already supported by digital tools.

One study in India showed that providing automated, personalized performance feedback to community health workers improved their self-reflection and increased their average number of client visits by over 20% [59, 57]. Another

RCT in Mali showed that using personalized analytics dashboards during face-to-face supervision of CHWs increased CHW activity by an average of 40 house visits per month, without compromising the quality or speed of care [262]. Yet these large scale and effective performance management systems typically do not systematically incorporate feedback from beneficiaries. Recognizing this shortcoming, we began our study with a particular interest in connecting people who receive care to the feedback systems that already exist within the health sector. Before considering how we approached this matter empirically, we would like to review one more body of related work that informed our exploration of feedback practices.

3.2.3 Socio-Cultural Challenges with Feedback

Collecting critical feedback from communities in low-resource contexts is often challenging due to social and cultural differences between researchers and their participants [3, 101, 108, 146]. Particularly when users in underserved communities are asked to provide feedback about artifacts, they are often biased because they worry that critical feedback could negatively impact their relationships with organizations that support key services [3, 101]. Dell et al. [56] demonstrated that participants in India were 2.5 times more likely to choose a technological artifact that they believed was developed by a researcher, even when the alternative was identical. While often discussed as a methodological concern in ICT4D research, response bias could surface as a design challenge in any attempt to construct feedback systems in these settings.

To mitigate response bias, ICTD researchers have applied social proofing, a psychological construct, to surface critical feedback that may be relevant for improving an ongoing project. For example, letting users believe that others in their neighborhoods provided critical feedback may encourage them to act similarly [255]. Other techniques include spending more time with participants in the hope that they eventually will become comfortable enough to provide critical feedback [76], adopting dramatic storylines in user studies [37], and role playing with skits [172]. While broadly relevant to research methods, some of these strategies are more amenable than others to inform the design of routine feedback systems. In light of this work, one of the goals of our study was to identify which strategies seemed most relevant and to imagine the specific ways that they might be used in system design efforts.

3.2.4 Designing Universally Accessible Tools

A wealth of literature has examined the design of tools that use only the functionality provided by basic mobile phones (i.e., not smartphones), such as voice calls and short message service (SMS) [192, 256, 58]. However, one universal communication channel that remains relatively unexplored is USSD (Unstructured Supplementary Service Data). Also called short-codes or quick-codes, USSD is commonly used for carrier service requests (e.g. users dial *144# to check an airtime balance with Safaricom in Kenya) and mobile money transactions (e.g. Telenor's Easypaisa service in Pakistan). USSD has a number of advantages over SMS. For example, USSD preserves user privacy as interactions leave no visible traces on the device, and its user interface dialog visually lets a user know when an interaction begins and ends. Also, in contrast to SMS, which is an asynchronous communication channel, USSD supports stateful, synchronous communication, and its interactions happen in real time [193]. Researchers have studied USSD in mobile money interventions [197] and information collection [260]. However, prior work on USSD is very limited compared to SMS and voice-based systems. Our works expands this literature by exploring a USSD system that collects feedback from care recipients about community health services in rural Kenya.

3.3 Part I: Qualitative Research on Identifying Stakeholders Groups and Current Feedback Practices

Our research began with a qualitative study to identify the diverse groups of stakeholders in the health ecosystem. We also wanted to find out current feedback practices including how feedback is collected from care recipients, how it is used, and the challenges involved in this process. Then we built on the insights gathered to create a digital system for care recipients to share feedback about the health services they receive in their communities.

3.3.1 Research Context

Our IRB-approved qualitative research took place in Kenyenya, a district in Kenya. At our research site, frontline workers were referred to as *"Community Health Volunteer (CHV)"* but throughout this paper, we use *"Community Health Worker (CHW)"* as it is more familiar in the ICTD literature. We worked with two organizations: Living Goods, a non-profit organization that supports networks of 'Avon-like' entrepreneurs to sell essential household commodities and address child health, nutrition, and family planning needs door to door; and Medic Mobile, a non-profit tech company that designs and implements open source software for health workers in hard to reach communities. Both organizations work in partnership with the Kenya Ministry of Health. Since 2014, Living Goods has partnered with Medic Mobile to co-design a digital health system to support their CHWs and supervisors. The system includes the Smart Health mobile app for CHWs (all CHWs are equipped with Android smartphones), a supervisor mobile app, and a web dashboard for supervisors to monitor CHW performance (see Figure 3.1). Source code and more information about this system are available on Github[1]. Our study built on this collaboration and began as part of an ongoing human-centered design project, focused on beneficiary feedback and led by Medic Mobile's design team.

Our research context featured two types of supervisors: MoH supervisors from the government and Living Goods supervisors. All CHWs are employed by the Ministry of Health and partner organizations like Living Goods can recruit a subset of CHWs to receive additional training and provide additional services (e.g. doorstep treatment of pediatric malaria) that are beyond the scope of other government CHWs. As such, the CHWs we recruited from Living Goods reported separately to both MoH supervisors and their Living Goods supervisors. As a sustainability strategy, Living Goods did not provide free medications to beneficiaries and did not pay CHWs a regular stipend because CHWs could earn commissions from selling health products to beneficiaries. The dual aim of this strategy is for fully active CHWs to earn more in commissions than they would with a flat salary and asking beneficiaries to pay a nominal fee for drugs delivered to the household often costs less than patients would pay for transportation to a health clinic where medicines are provided free-of-charge.

3.3.2 Participant Characteristics

We recruited 23 participants (nine Living Goods supervisors, two MoH supervisors, seven CHWs, and five beneficiaries) through Living Goods' office in

---

[1] See https://github.com/medic

23 Participants	Beneficiaries (5), CHWs (7), Supervisors (11)
Age	Beneficiaries: Min (20), Max (60), Avg (32) CHW: Min (29), Max (51), Avg (40) Supervisors: Min (23), Max (37), Avg (28)
Gender	Beneficiaries: Female (5) CHW: Female (3), Male (4) Supervisors: Female (6), Male (5)
Education	Beneficiaries: form two - secondary school CHW: primary - secondary school Supervisors: diploma - masters
Occupation	Beneficiaries: farming (5) CHW: 1 - 18 yrs experience Supervisor: 11 months - 4 yrs experience
Phone	Beneficiaries: feature phone (5) CHW: feature phone (4), smartphone (3) Supervisor: smartphones (11)

Table 3.1: Participant demographic characteristics.

Kenyenya. A supervisor at the branch reached out to Living Goods and MoH supervisors, CHWs, and beneficiaries to invite them to participate in the study. Interested people then came forward and we did spontaneous interviews with them over one week with about two to four people daily. Participants were not compensated.

As shown in Table 4.1, beneficiaries were all female, ranged in age from 20 to 60 years (average = 32 years), had a minimum education level of form two, had farming as their occupation, and used feature phones. CHWs included three females and four males, ranged in age from 29 to 51 years (average = 40 years), had a minimum education level of primary school, with one to 18 years of experience as CHWs, used feature phones (4), and smartphones (3). Supervisors included six females and five males, ranged in age from 23 to 37 years (average = 28 years), earned a diploma as a minimum level of education, had eleven months to four years of experience as supervisors, and all had smartphones. Living Goods supervisors also had laptops available in their offices.

3.3.3 Qualitative Methods

We conducted eight focus groups with 20 participants in groups of two to four people: three groups of Living Goods supervisors, one group of MoH supervisors, two groups of CHWs, and two groups of beneficiaries. We also did three semi-structured interviews: the first one with a female beneficiary in her sixties (45 minutes), and the others with Living Goods supervisors (30 minutes and two hours).

We tailored our questions to the kinds of participants we interacted with. The discussions involving supervisors focused on their current feedback collection practices, the kinds of feedback they received, how it is used, the protocols they used when beneficiaries were directly called, and the challenges they encountered in the feedback process. With CHWs, we discussed their routines during household visits, how they collected feedback and shared with their supervisors, how frequently it was received, tensions related to cultural dynamics, and how Living Goods' policies affect them.

For beneficiaries, we inquired about the services they received, what they discussed with CHWs, how they dealt with issues, and the reasons behind their approach to resolving issues. Interviews were conducted by two members of the research team in both English and Kiswahili, with one co-author serving as a translator. In total, our data set consisted of 12 hours of focus groups and 3.25 hours of interviews.

3.3.4 Data Analysis

We audio-recorded and transcribed our interviews and focus groups. We then performed thematic analysis [234] on the transcripts and our field notes, beginning with a close reading of the transcripts and allowing codes to emerge from

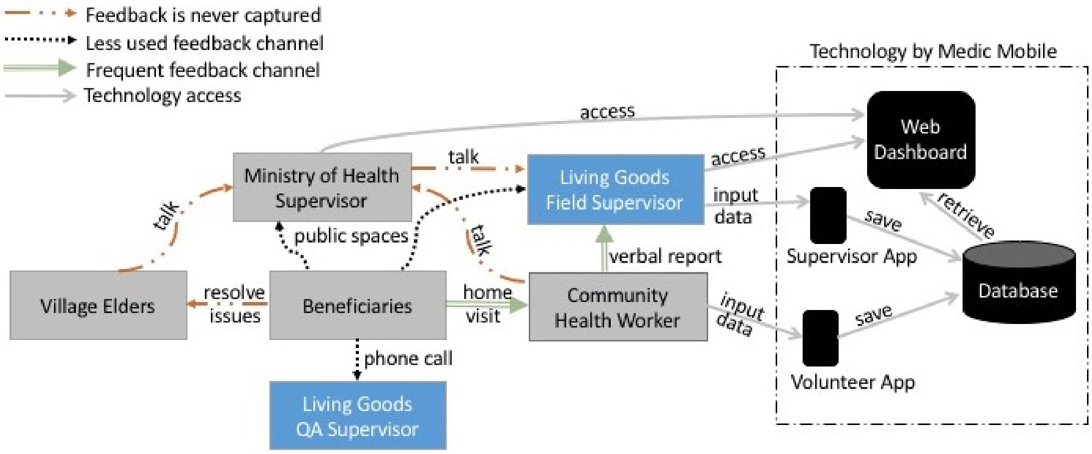


Figure 3.1: Summary of how beneficiary feedback is shared across diverse stakeholders. The blue-colored boxes are employees of Living Goods. Field supervisors directly interact with Community Health Workers (CHWs) while Quality Assurance (QA) supervisors call beneficiaries to verify that CHWs did household visits. The technology supporting feedback flow consists of mobile applications, a web dashboard, and a backend database.

the data. Multiple passes through the data resulted in 26 distinct codes (e.g., *feedback in public*, *beneficiary conflict*, and *overpriced medications*). We clustered related codes into high-level themes (e.g., *chance encounters*, *feedback improves services*, and *sustainability model*) and organized them in a codebook. After multiple discussions and iteratively refining the codes and themes, the research team arrived at a final set of themes that comprehensively represented the data.

3.4 Findings from Qualitative Research

Our findings reveal a deeper understanding of the environment in which feedback occurs. We uncover ongoing practices of how feedback is collected, the kinds of feedback collected, how feedback is used, and the socio-technical factors that impact feedback systems.

3.4.1 How is Feedback Currently Collected?

We begin our exploration of beneficiary feedback systems by examining how feedback is currently collected. Our data reveals three main ways in which Living Goods currently gathers feedback from beneficiaries: direct phone calls to households, indirect CHW reports that are sometimes shared with supervisors, and chance encounters in public places. We discuss each of these in turn.

Direct Phone Calls to Households

A number of supervisors (n=4) on the Quality Assurance and Control team explained how they place direct phone calls to households as a formal way to collect feedback from beneficiaries (see Figure 3.1). These phone calls primarily help supervisors confirm if CHWs truly visited households as recorded on the CHW app. One supervisor described:

> *"Every quarter, we randomize at least 30 visits or 30 registrations by a CHW per branch and then we have backend calls to the clients... We have a screening pool that we go through to see whether the services that [health workers] said were offered were actually offered, whether it's a true service, what [care recipients] thought of the quality of the service... and any other feedback they might have for us."* (P12, Female, QA Supervisor)

When supervisors call households, they follow an open script where they introduce themselves, assure confidentiality of the discussion, and ask if beneficiaries know any Living Goods CHWs. At the end of the call they ask for a 1-10 Likert scale rating of the services beneficiaries received and rationale for their rating. Although these direct-to-household phone calls are the most formal method of collecting beneficiary feedback, our data shows that this results in a range of challenges.

One major challenge is the small number of beneficiaries reached by the process. Beneficiaries eligible to receive calls are limited to those that provided phone numbers (roughly 65% of beneficiaries), and then by those who were visited by a CHW in the last quarter. Then, of these filtered lists, only 30 beneficiaries are chosen to receive calls. This is because the process of making individual calls to beneficiaries is time consuming and resource intensive. Supervisors explained that it was common for beneficiaries to discuss non-health related issues on these calls and, to remain courteous, supervisors listened to whatever the beneficiaries wanted to discuss. Although this may build rapport between Living Goods and the community, it is not an efficient way of collecting feedback. It is also common for phone calls to not be answered. The small number of beneficiaries reached is problematic for several reasons. From an organizational perspective, Living Goods may not receive sufficient information to be able to evaluate their services and products. In addition, the poorest households are less likely to own phones; the fact that they are excluded from this process makes the feedback system less equitable.

Direct-to-household phone calls may also be problematic for situations in which families share a device. In particular, it is common for a household member who is not the beneficiary to answer the call, which leads to privacy and confidentiality challenges. As one supervisor explained,

> *"A [teenager] gave us her father's number during CHW registration so when the supervisor calls that number saying 'your daughter is pregnant,' he says 'no, my daughter is not pregnant... yes, that is my daughter but she's not pregnant because I don't know [about it]... I live in Akulo and my family lives in Kisii'... you just asked the rightful owner of the phone, but he or she doesn't know all the information."* (P5, Female, Supervisor)

Another challenge supervisors (n=3) highlighted is that, due to the fragmented nature of the healthcare ecosystem, in which CHWs may have multiple affiliations (e.g., MoH and Living Goods), many beneficiaries who receive calls may not know or have forgotten what Living Goods is. A supervisor said,

> *"CHWs may forget to tell clients they are Living Goods CHWs, so the Quality Control team has to give many details to clients so they recall who Living Goods is... Some clients are agitated that you got their phone number . . . You are calling someone who could be having a thousand and one problems and yet you say "Hi I'm calling from Living Goods."* (P21, Female, QA Supervisor)

Finally, although in-person phone calls hypothetically provide opportunities for rich conversations with beneficiaries, the Quality Assurance supervisors (n=4) explained that the team is currently primarily focused on feedback that simply confirms whether CHWs did the work that they reported—visited households, provided treatment, and sold health products.

CHW Informal Report to Supervisors

Although direct-to-household phone calls are the most formal feedback mechanism currently used, they were not the most common method of reporting beneficiary feedback. Instead, we discovered that informal verbal reports that CHWs discuss with their supervisors are the most prevalent beneficiary feedback reporting mechanism. All supervisors (n=11) and CHWs (n=7) described how, when CHWs meet with their supervisors to go over their performance, CHWs volunteer feedback received from beneficiaries during their household visits. As one supervisor described,

> *"The CHWs kind of trust us that they will tell us everything that happens in the community even things that don't involve Living Goods."* (P2, Female, Supervisor)

However, this mode of collecting and reporting feedback is not mandatory and is not done in a systematic or structured way. Further, the process is complicated by the fragmented nature of the ecosystem. As previously described, Living Goods CHWs have (at least) two supervisors: one from the MoH and another from Living Goods. As a result, CHWs may sometimes discuss beneficiaries' feedback with their MoH supervisor and other times with their Living Goods supervisor, but there is no systematic way of capturing the information shared. Moreover, supervisors told us that they are only interested in matters that pertain to them. For example, Living Goods supervisors only want to talk about issues that pertain to Living Goods, such as pregnant women using medications, and are not necessarily interested in matters that pertain to the MoH, such as households refusing to purchase toilets because they engage in open defecation.

In addition to lack of structure, relying on CHWs to relay beneficiaries' feedback to supervisors clearly suffers from a number of biases, including recall bias [45], a psychological phenomenon where one inaccurately recalls past experiences. Even if a CHW wanted to share full details of the feedback received from a beneficiary, they may forget some of the details due to the time lapse between meeting with the beneficiary and their supervisor (CHWs meet with their Living Goods supervisors a few times a month).

Perhaps more importantly, relying on CHWs to report beneficiaries' feedback may bias the process towards only collecting positive feedback, partly because beneficiaries are unlikely to report negative feedback to their CHW for fear of causing problems. As one beneficiary described,

> *"I will just stop seeking their service. I will not tell anyone because there is no one to tell about it. If you start bad-mouthing the CHWs, it'd bring bad reputation to Living Goods . . . involving others brings about unnecessary friction . . . I'd rather finish with this CHW and find another CHW . . . I will go to another one to ask for the services I need but I will not tell why I am seeking out a different CHW."* (P18, Female, Beneficiary)

Although switching CHWs may enable beneficiaries to seek services from a different CHW, it does not necessarily provide them with a safe channel for providing negative feedback, since beneficiaries worry that CHWs know and will communicate with each other. One beneficiary told us,

> *"You might go and say something to another CHW but you don't know their relationship with the previous CHW you are complaining about [their services]... so it means you have started something you shouldn't have."*
> (P17, Female, Beneficiary)

Finally, even if beneficiaries feel comfortable telling their CHWs their negative feedback, it is unlikely that the CHW will relay this information to their supervisor, especially if such feedback reflects poorly on the CHW. By contrast, CHWs are more likely to report feedback and stories that reflect positively on their work. We provide concrete examples of both positive and negative feedback later in the paper.

Chance Encounters and Public Events

A third method currently used to collect beneficiary feedback is through chance encounters between supervisors and community members in public spaces, and during public community events. Half of the supervisors (n=7) explained that it was common for beneficiaries to approach them in public places and provide feedback to them. Two beneficiaries also explained how they identified supervisors walking around in their communities and spoke to them. One shared,

> *"I could report a CHW to supervisor when I see the supervisor in the community and I am comfortable talking to him... I have talked to a supervisor in the past when I saw a group of supervisors walking around in the community... if a supervisor is not around I will go to the hospital. I cannot go to Living Goods office to talk to supervisors because it is too far."* (P20, Female, Beneficiary)

Living Goods supervisors can be easily identified in public places through their uniforms while MoH supervisors are well known by the village elders of communities. Some participants (two supervisors, two beneficiaries) explained that they had observed others provide feedback during a public event. One supervisor told us: *"Sometimes they do announce it at the barazar [public community meeting] and make it public"* (P6, Male, Supervisor).

Beneficiaries had mixed feelings about using community events and market places as an avenue for connecting with supervisors and providing direct feedback. Once again, we saw that this method of providing feedback is more conducive to receiving positive feedback, with a supervisor commenting that the community *"can say that the people you have given us are doing a good job"* (P7, Male, Supervisor). However, three beneficiaries said they would not share negative feedback during public events and do not consider it wise for others to do so for fear of escalating an issue. A beneficiary said,

> *"It is a bad idea to resolve an issue at a barazar [public community meeting]. There are other ways that are better...you can tell the CHW yourself. If in a*

> *situation where it is really serious, you can go to the village elders to help solve the issue."* (P18, Female, Beneficiary)

All beneficiaries felt that telling someone else about challenges they faced with their CHWs may be construed as spreading gossip. As such, they preferred to remain quiet about CHW issues they encountered. One beneficiary shared,

> *"I will not tell anyone because I don't like to gossip. I will tell my husband and just endure it... If it is too trivial to tell my husband I will not tell him because he might go fight about it and escalate it... I am not very confrontational and don't want there to be pain and for things to escalate. I'd rather talk to the CHW I disagree with instead of making it escalate by involving others."*
>
> (P17, Female, Beneficiary)

In addition to chance encounters in public, four participants (two beneficiaries and two supervisors) told us that beneficiaries talk with village elders when they had issues with CHWs, and preferred to resolve issues locally instead of involving supervisors from MoH or Living Goods (see Figure 3.1). One beneficiary described,

> *"If you're in a situation where it is really serious, you can go to the village elders to help solve the issue."*
>
> (P18, Female, Beneficiary)

Village elders in turn provide feedback to supervisors when they see them in their communities. It is possible that these elders interact with supervisors from partner organizations, but our data suggest that they mainly interact with MoH supervisors (see Figure 3.1). As such, their feedback often does not make it to Living Goods. One MoH supervisor shared,

> *"So they know if they go to me with certain issues about the CHWs that I will be able to make a decision."*
> (P7, Male, Supervisor)

The two MoH supervisors we interviewed explained that MoH supervisors can investigate reported issues by presenting the issue to an internal committee, who has the authority to take disciplinary actions such as terminating a CHW's job.

3.4.2 What Kinds of Feedback are Collected?

At a high-level, we can separate feedback into positive feedback praising the CHW and/or organization, and negative feedback that reports issues or complaints with services received.

All Living Goods supervisors (n=9) described how they received a lot of positive feedback from communities expressing their happiness and satisfaction with the CHWs' work and the effectiveness of the products, such as malaria medications. One supervisor explained,

> *"Sometime late last year, there's this mother that her kids had malaria and it kept recurring until this CHW went in at the middle of the night and treated this kid without the mother leaving the house and she was really appreciative like 'we thank you so much for bringing the CHWs on the ground. They really help us... My kid took the malaria medication after the CHW tested and the second day my child was able to play."'* (P1, Female, Supervisor)

One common source of appreciation stemmed from the fact that CHWs visited beneficiaries at home so that they did not need to stand in queues at their local hospitals. One shared,

> *"Government facilities have long lines so it is just easier for me to call the CHW. Otherwise it is a waste of my time... I need to go on a motorbike to get to the hospital...I can identify a good CHW as one taking time out of their day to come see me."* (P18, Female, Beneficiary)

Many participants (five supervisors and four beneficiaries) shared examples of what they consider to be *"good"* CHWs performance in which CHWs helped community members in dire health situations or went above and beyond their daily responsibilities of performing two hours of household visitation. One beneficiary shared,

> *"The CHW found out I was pregnant and he sent me messages on how to eat, how to take care of myself, up to the moment I delivered my baby and even after my delivery. That CHW was really helpful."* (P18, Female, Beneficiary)

Compared to the abundant examples of positive feedback, participants provided only a few concrete instances of negative feedback. We also noticed that, after describing a few examples of negative feedback (provided below), many participants turned to hypothetical instances of negative performance. For example, care recipients would report negative feedback as *"I heard this from someone else"* instead of *"this happened to me"*, and when we followed up about the impact it had on them, they told us, *"it did not happen to me"* .

Supervisors explained that, although infrequent, beneficiaries sometimes provide negative feedback during direct phone calls about CHW performance, such as *"the CHW is not working well"*. One relatively common source of negative feedback (brought up by three supervisors and all beneficiaries) occurs when CHWs become unreachable when they are needed, leading to perceptions that CHWs are not invested in their job. One beneficiary shared,

> *"The [CHW] does not respond when you call and acts as if they are being pushed to do the job."* (P18, Female, Beneficiary)

Another frequent (five supervisors, all CHWs, and four beneficiaries) negative issue that came up relates to Living Goods' sustainability model, in which CHWs sell medications to beneficiaries. Since several other programs have provided medications for free, beneficiaries were often not happy when asked to pay. We discuss this and other issues related to the sustainability model later in the paper.

3.4.3 How is Feedback Currently Used?

Our data shows that feedback collected from communities provides diverse benefits to Living Goods by helping them understand the impact of their services, improve training, motivate CHWs, and detect fraudulent behavior.

Many supervisors (n=10) explained that insights from beneficiary feedback are used to improve CHW training and delivery of services. Relevant information gained from direct phone calls to households are passed on to field supervisors which may come up as a topic to address in CHW-supervisor meetings and for other Living Goods departments, such as the marketing team. One supervisor explained,

> *"The feedback we receive is used towards improving services . . . For example, telling us about our products and saying that it was expensive, this information goes to the marketing department. Some provide feedback saying that our CHW is not efficient and this can help us in the process of CHW training. When we identify negative feedback . . . we ask the manager to focus more on it . . . and then closely monitor the CHW to see if they could be further trained."* (P21, Female, QA Supervisor)

Feedback is also important in helping Living Goods understand if CHWs adhere to the care protocols and procedures for which they have been trained and ensure that CHWs do not overstep their level of expertise. One supervisor shared:

> *"[Beneficiary feedback] helps us to understand if [CHWs] treat what we focus on, or go there and overdo to say [they] treat all the diseases. The information the [beneficiary] gives us helps us to assess that. We see the weaknesses and it helps us to focus on those during the [CHW-supervisor] monthly meetings so that it is clear to [the CHWs] and so that it doesn't happen again."*
> (P2, Female, Supervisor)

Supervisors and CHWs felt that positive feedback acted as a driving force that motivated CHWs to keep serving their communities. Every month all CHWs and supervisors in a district meet as a group where they discuss issues and share positive feedback from their communities. This feedback primarily focuses on success stories from community members. A supervisor explained,

> *"The feedback goes a long way for us, we feel like we changed somebody's life... we've changed the mentality of how people viewed [their health services]... If a CHW has a success story, I tell them to share by word of mouth during our monthly in-service meeting...we share with the rest of the CHWs."* (P1, Female, Supervisor)

Another prominent use of beneficiary feedback is to help detect fraudulent CHW behavior and/or data fabrication. Some supervisors (n=3) explained that CHWs could fabricate the number of home visits in their mobile app. As such, supervisors are always on the lookout for fraudulent behaviors, asking beneficiaries when they accompany CHWs on home visits or via in-person phone calls to households. A supervisor shared,

> *"There were times we used to get fraudulent data about clients and we have reduced it. When we call beneficiaries, they are able to give us the actual data... So we see that some of our CHWs who had high performance [that were outliers] now came back to normal... We have been able to improve our services through coaching [CHWs] because our [beneficiary responses] helped us to find [fraudulent] data."* (P21, Female, QA Supervisor)

Finally, supervisors explained that receiving beneficiary feedback has helped them revise their key performance indicators for CHWs. For example, in the past, high levels of fraudulent activity were detected among CHWs after they were told they would receive monetary incentives if their performance increased. When responses from beneficiaries showed that CHWs had been fabricating their records, the organization changed its key performance indicators to instead focus on the quality of data reported by CHWs instead of increased performance. This example illustrates the power of beneficiary feedback to impact organizational work practices.

3.4.4 What Socio-technical Factors Impact Feedback Systems?

Having described how feedback is collected and used, we now describe some of the socio-technical complexities that impact the environment where feedback is collected and the tensions that arise between stakeholders.

Community Relationships

Since CHWs are chosen from and embedded within the communities that they serve, they often have preexisting relationships with different people in the community that may impact their work. In some cases, these relationships might be cordial leading to positive feedback. In other cases, they could be negative regardless of how well CHWs carry out their work. As one supervisor shared,

> *"There are some CHWs who relate well with their communities. Some of them are family members and their community really loves them. So for those CHWs, we get a lot of positive feedback. Their relationship is different compared to other CHWs even though the other CHWs are doing a really good job."* (P1, Female, Supervisor)

Moreover, since CHWs are frequently collecting sensitive health information from beneficiaries, the nature of their relationship with the beneficiary may affect how they do their work. For example, several supervisors (n=8) described how community members sometimes feel CHWs are asking for information that is too personal. One supervisor shared,

> *"Most of the information that the client gives our CHWs is personal information. Like when registering a pregnancy, 'when was your last menstrual period?' This is a male CHW, this is a mother you are asking. This is not your wife, this is not your relative. So some of the clients don't give actual information because they are like 'why is he asking me how many children I have? If I have ever had a miscarriage?' but in pregnancy care, you have to know these things so that you see if there's a risk factor."* (P1, Female, Supervisor)

Half of the supervisors (n=6) told us that they encouraged CHWs to maintain strong relationships with their communities by treating beneficiaries when possible, regardless of who the beneficiary is. A number of CHWs (n=4) told us that it was common for them to treat community members outside their officially assigned households (each CHW is assigned 30 to 100 specific households).

However, treating beneficiaries outside a CHW's area, or *"crossing boundaries"*, increases the complexity of collecting feedback. CHWs explained that when they treated beneficiaries outside their boundaries, it was not reflected in their CHW app. One CHW told us,

> *"I treat them, ask the other CHW to record the community member's data and then ask that CHW to give me back the medicine I gave out in order to make sure that the stocks count."* (P10, Male, CHW)

Since these out-of-bounds beneficiaries are considered to be outside the CHWs area and the treatment is not reported in their mobile application, CHWs often do not share any feedback that was received. As such, beneficiary feedback is lost in transmission during exchange of reports.

Beyond beneficiaries, CHWs also face challenges in their communities due to relationships with village elders. Supervisors (n=3) explained how the village elders may have issues with CHWs due to perceived power dynamics. One explained,

> *"Sometimes the village elders disagree with the CHWs. Maybe they think that the CHWs are being paid some little bit of money. The village elders are in charge of a certain village and the CHWs are also put in charge of households in the same village based on certain health indicators. Now the misunderstanding comes from the village elders. When they see the CHWs walking around with this air of jurisdiction they think that they [CHWs] are over-doing their work and that is when they [village elders] come up with [negative] issues about the CHW.'* (P6, Male, Supervisor)

Since village elders who are in conflict with CHWs may provide unwarranted negative feedback to supervisors regarding the CHW's performance, supervisors explained that they need to dig deeper and properly investigate any issues raised instead of taking it at face value. One supervisor shared,

> *"So sometimes we take time before we make a decision. You cannot rush and make a decision based on what the village elders say."* (P7, Male, Supervisor)

Sustainability Challenges

The majority of the negative feedback that Living Goods receives relates directly to their chosen sustainability model. As described earlier in the paper, Living Goods tries to achieve longterm sustainability by not paying stipends to CHWs, requiring CHWs to instead sell medications to community members instead of providing them for free. One supervisor shared,

> *"We don't pay stipends to [CHWs] because it is not sustainable. We have seen partners come on the ground and leave after two years but for Living Goods, we are here for the longterm so we have to look for a sustainable way of making sure that we are still supporting the community and not run out of medicines... Our prices are a little bit below or at the market level and the quality is so high."* (P1, Female, Supervisor)

CHWs are able to make commissions that range from 60 shillings ($0.60) to 600 shillings ($6.00) depending on the item sold[2]. A supervisor told us, *"the higher the price, the bigger the margin."* However, many participants (nine supervisors, seven CHWs and four beneficiaries) explained how the adopted sustainability approach has led to conflicts between CHWs and beneficiaries, which drives negative feedback on pricing. One participant described,

---

[2] For perspective, 120 shillings can purchase a meal in a local restaurant and 200 shillings can be used for a 20-minute taxi ride.

> *"I don't want the CHW to inflate the price. That's something I thought they could do... They say to me: 'This medicine is 50 shillings, can you add 10 shillings so I can eat a banana?'"* (P20, Female, Beneficiary)

Tensions surrounding money may affect both the services that beneficiaries receive and the process of collecting feedback. For example, the majority of our beneficiaries (n=4) explained that they would avoid seeing a CHW if the person asked them for extra money. One beneficiary shared,

> *"A CHW comes to me and he'd try to sell the medicine to me as a business saying 'I traveled to get here so instead of 100 shillings, I'll sell to you for 150 shillings.' If a CHW tries to increase the price of the medicine, I could just tell them no and avoid the service with them."* (P18, Female, Beneficiary)

For their part, the CHWs faced numerous challenges trying to negotiate money with their communities. One CHW told us how he wanted to preserve good relationships with the community and also provide people with necessary health services, so would often just ask community members to pay whatever they could. He said,

> *"I don't want people to run away from me so I ask them to give me whatever they can and I top up the rest myself."*
> (P8, Male, CHW)

CHWs also felt conflicted selling medications to beneficiaries they thought could not afford to pay and sometimes they provided the medications without asking for money. One CHW explained,

> *"Sometimes we don't get the profit. You just give because she is totally poor that you don't even want to sell to make profit... and the child is nearly dying. What can you do?"* (P11, Female, CHW)

Another common scenario we encountered was for a CHW to provide a community member medicines on credit. However, when CHWs went back to ask for payment, beneficiaries frequently refused to pay because they felt the medications should have been provided for free. All CHWs we spoke with were currently owed money by various community members (by two to five people). One CHW shared,

> *"These people believe that we were given the drugs for free. You just give drugs because you can't leave a child dying... When you go back for the payment, they say we are being given free... so it makes us to go to our pockets again to support the community."* (P9, Male, CHW)

At other times, CHWs completely avoid visiting households with a history of not paying for medications or health products. They also explained to us how, to avoid conflict with beneficiaries with delinquent debt, they secretly only gave items on credit to people they felt would pay. One CHW explained,

> *"You have to avoid people with bad history of payment. If two beneficiaries are here and one asks me to provide goods to be paid for later but the other is someone who does not pay and is watching my response, I'll quietly tell the first person that asked me to go and come back later so I can keep it secret that I gave the item on credit."* (P14, Female, CHW)

Unfortunately, if both CHWs and beneficiaries avoid each other due to tensions surrounding money, then the health services that the community receives and the feedback collected about those services will be negatively impacted. Interestingly, although about half of the CHWs (n=3) we spoke with suggested the need to pay CHWs a regular stipend, none of our participants brought up the need to increase community awareness of Living Goods' sustainability practices so that beneficiaries are aware they need to pay for medicines and know that CHWs are not extorting them.

3.5 Discussion of Qualitative Research

Having developed a nuanced understanding of how feedback is collected and used in the backdrop of socio-technical challenges, we now synthesize our findings into design opportunities for ICTD researchers and practitioners interested in gathering and using feedback in ways that engage a wide range of stakeholders, including care-recipients. In our discussion, we adopt a reflective design approach [215] and critically think about practical ways to embrace and build on the nuances we found in current practices. In particular, we discuss missed opportunities to collect and use feedback in an equitable and systematic way, and we propose practical design opportunities that adopt the parts of existing practices that work well while augmenting other aspects that could be strengthened.

3.5.1 Equitable and Systematic Feedback

A key finding in our research is that beneficiaries without mobile phones are not well accounted for in current feedback practices. This surprised us, because we initially had understood feedback activities as mostly analog and did not anticipate how phone access might matter. When Living Goods supervisors call households, they randomly select and call 30 beneficiaries who: 1) were visited in the last quarter, and 2) have registered a phone number with their CHW. Typically, some of the selected individuals do not answer the call, at times because they share the phone with someone who lives or works elsewhere. As a result, beneficiaries without phone numbers are systematically ignored and their feedback can only be heard if they have chance encounters with supervisors in public or if CHWs volunteer reports on the feedback that was received during household visits. Beneficiaries need an equitable channel to ensure that organizational decisions are not based on the voice of only the people with better access to mobile devices. Designing a feedback system that ignores this problem could further marginalize beneficiaries who are too poor to afford mobile phones.

While household visits and public encounters hold potential to engage a wider cross-section of the population, these feedback channels are better suited to positive feedback than to negative feedback. Beneficiaries consistently explained that they struggle to provide negative feedback directly to CHWs, because of the potentially fraught relational dynamics of criticizing a neighbor who lives nearby. Even when beneficiaries communicate directly with supervisors, they are reluctant to offer frank criticisms because they are worried about the ramifications of their actions. For example, they fear being labeled as *"gossips"* in their communities. This finding suggests that beneficiaries are prone to response bias, a well-known problem in the ICTD literature [56, 101, 255]. Beyond making feedback practically possible, an effective feedback system would also need to normalize the activity of surfacing criticisms, mitigate response bias in beneficiaries, and proactively deal with the potential unintended consequences that people may encounter.

In our particular empirical context, beneficiaries, CHWs and supervisors reflected on a specific organizational policy as a way of illustrating the dynamics of community feedback. However sensible Living Goods' user-fee strategy may be from a sustainability perspective, it emerged as a challenge for CHWs. Prior studies have similarly found that the design of remuneration models for CHWs can disrupt how they carry out their work [9, 220]. Part of the challenge has to do with policies that are easily misunderstood by beneficiaries. This was the case, for example, when beneficiaries expressed concern that CHWs who sell medications (rather than giving them away free of charge) may be extorting them. This is a compelling example of how NGOs might use feedback systems to navigate important but difficult policy questions in cooperation with the communities they serve. In some cases, organizations may identify opportunities to adapt the policy to deal with unforeseen edge cases and unintended consequences, such as when CHWs give away items on credit or forgo collecting payment because they consider a specific household *"too poor"*. In other cases, organizations may discover the importance of communicating more widely the constraints or strategies that inform a certain policy, so that the policy is better understood among the beneficiary population.

3.5.2 Design Opportunities for Feedback

Without conducting this qualitative study, we could not have learned of the challenges of equity affecting beneficiaries without phones. One design opportunity has to do with building feedback systems into tools already used by health workers, in ways that could be accessed by beneficiaries that may or may not have phones of their own. For example, when beneficiaries provide positive feedback during household visits, CHWs could record this in their apps, and they might also record critical feedback about care received at clinics. Supervisors could use their mobile app to record feedback when they encounter beneficiaries at a market place, and organizations could also consider having supervisors visit select households without the CHW present. By formalizing feedback in ways that cater to people who own devices and those who don't, feedback collection practices could become more systematic and more equitable.

To alleviate the fear of backlash, collecting both positive and negative feedback with anonymity may be worth exploring. Research has shown that anonymity can increase self-disclosure and empower people to better express themselves [156, 22]. Anonymous feedback channels could be designed in a way that augments current practices, rather than replace them. Supervisors and CHWs could continue to use current feedback channels, with the understanding that this feedback tends to be positive and is highly motivating. In parallel, beneficiaries could use an anonymous channel to provide all types of feedback, without fear of facing a backlash and/or being labeled in the community as *"gossips"*.

Since beneficiaries primarily use feature phones to make phone calls and send SMS, we propose augmenting current feedback practices with accessible technical approaches previously described in the ICTD literature. DeRenzi et al. [59, 57] used an Interactive Voice Response (IVR) system to successfully provide feedback to CHWs with low-end mobile phones, and they found that supporting multiple interaction modalities is beneficial for engagement. Vashistha et al. [255] showed that using social proof, by telling someone what others are doing, can mitigate response bias [56] and improve critical feedback from underserved communities. Building on these findings, we propose a beneficiary feedback system that caters to feature phone users, provides multiple interaction modalities, integrates anonymity, and leverages social proof. For example, beneficiaries could receive a dedicated phone number to an IVR system where they could anonymously call or text to leave detailed reviews or ratings about services received. They could also receive encouraging IVR prompts and SMS that socially-proof them to leave honest feedback by telling them the number of people who have recently provided feedback. Beneficiaries without mobile devices could indicate during face-to-face encounters with supervisors that their feedback should be recorded anonymously. Then all feedback collected can be made available as summarized web dashboard analytics used by supervisors to inform CHW training and improve the quality of services. Taken together, these opportunities suggest a path forward for systematic and equitable feedback.

3.6 Part II: Design of a Health Feedback System Based on USSD

In this part, we explored the design of a system that enables individual care recipients, or beneficiaries, to provide feedback on the community health services they receive. Although the target users for the system are care recipients, our design process involved multiple stakeholder groups including care recipients, CHWs, supervisors, decision makers, and more.

Iterating with multiple groups of stakeholders nurtured a spirit of community collaboration as each group learned the other groups had shared their views on the importance of a tool for care recipients.

We worked in Siaya, a rural area in southwest Kenya with a population of about 850,000 people. Communities in Siaya primarily receive health services in two ways: (1) when they visit a health facility; and (2) when CHWs conduct household visits (e.g. to check on a pregnant woman or provide medication). Our fieldwork team consisted of two women from East Africa and one man from West Africa who all had experience working in Africa. Two team members were very familiar with Siaya: one spent the previous year in Siaya cultivating relationships with stakeholders; the other had previously worked in Siaya. All research activities were approved by our university's IRB and by local Ministry of Health authorities (Director of Siaya County and other County leadership).

3.6.1 Design Goals and Challenges

Our design goals focused on creating a tool that takes into consideration the needs of all stakeholders, including care recipients and decision-makers (e.g., government or organization leadership). We focused on two contexts for care recipients to provide feedback: (i) during a hospital visit; and (ii) after receiving care from a CHW who visited a household. Throughout this chapter, we refer to the avenue for providing feedback about visits to health facilities as the *Hospital Line* and for CHW household visits as the *Household Line*. The initial goals and challenges described below were informed by prior research on community health feedback tools aimed at care recipients [181], design recommendations from experienced stakeholders who worked closely with our target users, and our prior work on cultivating a human-centered design process [103].

Access and Equity. Providing access is a critical part of designing for low-resource settings. Many technical tools for underserved communities build on universal communication channels available on any basic phone, including phone calls [256], SMS [192] and USSD [193]. Communities are often familiar with these channels, which reduces the training required to deploy applications. Users also do not have to install custom software to use tools implemented via these channels. Our work uses one such universal channel: USSD. It is also important that our tool work in areas with relatively poor connectivity. Similar to phone calls and SMS, USSD does not require Internet connectivity to work, thereby making our tool available to any mobile phone user in Kenya with cellular coverage.

Promoting equity works in tandem with providing access. For example, many tools deployed in prior research in low-resource settings provide free access, since charging users discourages engagement [250]. In addition, prior work recommended that feedback tools for underserved communities should not be limited to only a few community members because the feedback will not be representative of the community and may further marginalize vulnerable populations [181]. We chose to make our feedback tool free to use so it is accessible to as many users as possible.

Privacy and Anonymity. Prior work on CHW programs suggests that feedback tools need to preserve the privacy of care recipients to protect them from potential backlash by CHWs or others who may not be happy with the feedback [181]. In addition, research has shown that providing users with anonymity can increase self-disclosure and empower people to better express themselves [156, 22]. These design recommendations reinforced our decision to use USSD instead of creating an SMS or phone-call based system, since both SMS logs and call logs would reveal a person's usage of the system. By contrast, USSD interactions immediately disappear the moment a user completes their interaction with the system and leave no visible traces behind.

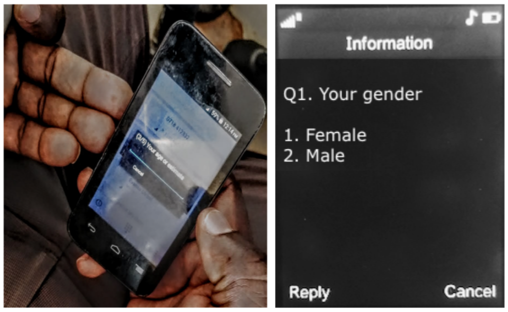
Generalizability. Another key design goal is collecting feedback about health services in a way that is useful to decision makers in community health programs. We sought to use standardized feedback metrics that fit our research context in Siaya while also serving as a relevant metric on a global scale, so that the system might be generalized to other contexts relatively easily. To achieve this, we chose to use the Net Promoter Score (NPS) metric [205] (*On a scale of 0 to 10, how likely is it that you would recommend [company, product, or service] to your friends or colleagues?*). This is widely used across global industries [206] and has been adapted into nonprofit settings [92, 128, 51].

Sustainability and Affordability. Another important design goal was to create a tool that decision makers could afford and potentially sustain over a long period of time. In Kenya, deploying a free USSD-based system is about two times the cost of creating an interactive SMS-based system for collecting feedback (i.e. about $500 more per year). However, deploying a phone-call based system is over 10 times more expensive when compared to a USSD system. Moreover, setting up an IVR system requires multi-weeks steps with official regulatory authorities in Kenya unlike SMS and USSD systems that could be approved in a few days. These financial projections further reinforced our decision to use USSD to create a feedback tool in place of adopting a phone-call based system.

Behavioral Design. At the core of our intervention is the intent to change people's behaviors on how they share feedback after receiving health services. Therefore, we engaged with Behavioral Science literature to inform our intervention design. We began this process by performing a review of behavioral techniques used in existing feedback systems around the world and assessed approximately 1100 documents (including peer-reviewed research and organizational reports). During our review, we focused on behavioral techniques that could be implemented and scaled through technology and arrived at three dominant techniques used in digital feedback systems: *positive reinforcement* [222], *framing effect* [245], and *social proof* [217]. However, we adopted only *positive reinforcement* because it did not require building up on a user's historical data like the other two concepts. Further, during early phases of discussing the purpose of our tool with CHWs and supervisors, they expressed concerns that care recipients felt the tool would be challenging to use. As a result, we focused on how to make users perceive the tool as one that is easy to use. We also reflected on ways to make the tool simplify users' decision making during the process of sharing feedback. As a result, we built on the theory of *choice architecture* [11, 117] to inform the placement of contents in our system because research shows that the order of presentation of items impacts people's decision making. For example, our tool showed users simple questions first (e.g. your gender) before asking them to later reflect on the kind of feedback they wanted to share. By interacting with the easy questions first, users would feel that the tool was indeed easy to use.

3.6.2 Implementation Details

Based on these design goals and prior work [181], we created an initial USSD prototype that used Africa's Talking [238], a third party service provider based in Kenya. The system was made available to all users on Safaricom, the most dominant mobile network in Kenya [47]. We implemented the system as open-source using Django web framework with an API endpoint for receiving USSD requests from users and a Postgres database deployed on Amazon Web Services. When a user dials the USSD code, it is converted to a network request by Safaricom, which redirects the request to Africa's Talking, who forwards the request to our web application. The web app generates a formatted response that is sent back to Africa's Talking, forwarded to Safaricom, and rendered on the user's mobile phone. This process happens for each question until the network forwards an "END" signal that originates from the web app after the last question has been



Q1. Your gender

1. Female
2. Male

Hospital Line (10 Hops)	Household Line (8 Hops)
1. USSD initialization (dial in)	1. USSD initialization (dial in)
2. Choose language	2. Choose language
3. Introduction – continue/exit	3. Introduction – continue/exit
4. Your gender (Q1)	4. Your gender (Q1)
5. Your age (Q2)	5. Your age (Q2)
6. Select the hospital (Q3)	6. NPS question (Q3)
7. Select the department (Q4)	7. Your reason (Q4)
8. NPS question (Q5)	8. End: Thank you
9. Your reason (Q6)	
10. End: Thank you	

Figure 3.2: On the left, a user dials into the system and answers a question about gender as shown in the next image. Completing all steps in the *Hospital* line involves 10 hops while the *Household* line involves eight hops.

answered.

3.6.3 Iterative Design through Fieldwork

After creating an initial prototype, we conducted iterative design with five stakeholder groups: care recipients, CHWs, supervisors, decision makers (county and sub-county leadership), and professionals at Medic Mobile (an NGO we worked with). In total, we did 7 rounds of iteration with 45 participants: 19 care recipients, 7 CHWs, 3 supervisors, 4 hospital leaders, and 12 NGO staff.

We began by explaining the goal of the system. Then we asked participants to tell us what mobile networks they used on their device, if they knew how to check the airtime balance on their devices, and how they did it (checking airtime is one of the common use cases of USSD). This gave us a sense of how many participants were familiar with USSD and could test on their personal devices. We also bought a basic phone with an active Safaricom number in case participants did not have a mobile device or a Safaricom line. Two of the 45 participants in our design process did not have a mobile device (both care recipients). Everyone who did have a phone also had a Safaricom line.

We asked participants (except care recipients) to imagine a scenario from the perspective of a care recipient. For example, for hospital feedback, participants imagined completing a hospital visit before dialing the USSD code. After completing the scenario on providing feedback, participants shared how they felt about using the USSD system, the questions they had answered, and challenges they encountered or they anticipated care recipients would face. We also followed up on any observations we had that they did not bring up (e.g. getting stuck on a question). For decision makers, we asked the ways in which the system could be integrated into existing workflows at the hospitals. We used a similar procedure to test the system for collecting feedback about CHW home visits, asking participants to imagine themselves as care recipients who had been visited by a CHW at their home. Figure 3.2 shows examples of USSD questions and a user submitting feedback.

We tested with 14 care recipients at hospital facilities and five at their homes. At the hospital, we approached people at the exit gates and waiting areas, spontaneously asking them if they were willing to spend a few minutes to test a system. If they agreed, we shared the goal of the project and asked those who had their mobile devices to test the system by dialing in using their own phones. If they preferred, we provided them with a basic phone that we carried with us. We followed a similar procedure for testing with people at their homes.

After each round of testing, the research team met and combined notes. We also analyzed the system logs during and after testing to understand the interactions that occurred (e.g. how long users spent on each screen). Based on these insights, we decided on any changes that should be made before further testing. We now discuss lessons learned through our design iterations.

3.6.4 Findings and Lessons from Design Iterations

Navigate mobile network constraints. A recurring theme in our fieldwork was that users who spent more than one minute on a USSD screen (single question) received a timeout error that was automatically generated by the mobile network, which required the user to restart the submission from scratch. For example, in an early version of the system, participants frequently timed out before they could type a response to *"Any other comment?"* To combat this issue, we changed our design so that all questions only required numeric input (rather than text). We also ideated about how we might inform users at the beginning of a USSD session that questions should be answered as quickly as possible, but ultimately decided against this since it might compromise the quality or accuracy of feedback submitted. Interestingly, our design iterations did reveal that many users were familiar with timeout errors, and usually redialed into the system when they happened.

Another constraint was that each USSD screen could have at most 154 characters (including invisible whitespace). Any text exceeding this length led to a confusing user experience where the extra characters were split into a next screen that could only be viewed by typing '98' or '0' to go back. As a result, we worked to design each of our USSD screens to have fewer than 154 characters, which made it challenging to balance creating questions that were long enough to not lose meaning after translations yet short enough to not be automatically split into other screens.

Shorten codes. Our prototype used the codes **384*11100#* for the *Hospital Line* to collect feedback about health facilities and **384*99900#* as the *Household Line* for feedback about CHWs who visited care recipients in their homes. We used two different codes to separate responses for the two contexts. However, testing with stakeholders quickly yielded suggestions that we shorten the codes since community members are used to dialing shorter codes (e.g. **144#* to check airtime balance on Safaricom). Thus, we worked with the USSD service provider to find shorter codes, specifically **384*777#* for the *Hospital Line* and **384*888#* for the *Household Line*. We were unable to make the codes shorter because (i) setting up a 3-digit code is over 10 times more expensive than setting up a 6-digit code; (ii) it takes multiple weeks to get approval from the telecom networks for 3-digit codes compared to a few hours for 6-digit codes; and (iii) it is non-trivial to find an available 3-digit code that will not be confused with other USSD services in Kenya.

Balance stakeholder suggestions. We frequently received conflicting suggestions from stakeholders. For example, CHWs were interested in knowing which community members submitted feedback about them, conflicting with care recipients desire for anonymity. As another example, decision makers often asked us to include lengthy and complex questionnaires, which would have been very cumbersome via USSD. Although we did not incorporate every stakeholder suggestion, we did manage to include some suggestions, such as adding two questions to the *Hospital Line* that asked care recipients to specify the specific facility and health department they visited.

Adapt standardized questions. Feedback from CHWs and care recipients revealed that the initial feedback question based on the Net Promoter Score [51] was challenging for users to fully comprehend. For example, asking *"How likely are you to recommend this hospital to someone else? Use a scale of 0 (highly unlikely)* *to 10 (highly likely)"* resulted in users selecting only 0 or 10 because they did not understand what the numbers in between stood for. To address this problem, we used cultural analogies such as *"think of it as marks you earn in school and give a score where 0/10 means bad and 10/10 means excellent"*. However, users said that a *"score"* made them think of a soccer match and not a hospital or a CHW. After many iterations, we eventually changed *"the scale of 0 to 10"* to three options: *"definitely yes"*, *"maybe"*, and *"definitely no"*, which was well received.

Integrate language preferences. During our fieldwork, we asked participants their preferred language and received three answers: English, Kiswahili, and a local language in Siaya (Dholuo). There was no consensus, with different individuals preferring different languages. For example, some explained that they did not want the system in Kiswahili because *"it is not our mother tongue, we don't speak it."* To account for these varying preferences, we redesigned the system so that questions could be rendered in any of the three languages, but without any single question in any language exceeding 154 characters in length. This process was quite challenging, with translations of English words often being longer in the other two languages. In addition, Dholuo had varying dialects, so substantial effort was put into using words that were well known in the local dialect. We validated our translation with people from different communities to find the most common terms and tested the translations with CHWs and care recipients during our iterations.

3.7 Field Deployment

After converging on our final design, we deployed the tool for seven weeks in Siaya. We trained five CHWs to use the tool and they in turn trained community members to provide feedback in two contexts: at the hospital and during home visits. All CHWs trained care recipients during household visits, while only one CHW trained care recipients at a health facility. Each CHW reported being responsible for 50-100 households, making between three and 22 household visits per week. After two weeks, we held focus groups with stakeholders to understand their experiences using the system and adoption challenges. We now provide details of our deployment.

Procedure. We initiated our deployment by training five CHWs to use the system, and they in turn trained community members. We limited the training to only five CHWs because we wanted to get a sense of how the system might work at a relatively small scale and address any potential issues that arose before scaling to a larger group. We started by asking CHWs if they knew how to check their airtime balance, and everyone indicated that they were proficient with doing this. Then we discussed the potential to improve service delivery by collecting feedback, and introduced our USSD tool as a mechanism for care recipients to become connected to the health feedback loop. We then observed CHWs as they used the USSD system several times for each of the two codes: for hospital feedback and household feedback, and answered any questions or issues that came up. Finally, we asked CHWs if they felt comfortable teaching care recipients to use the tool, and ensured that they emphasize that it is free to use. During training, we were careful to explain that any feedback collected would not affect the CHWs' jobs and could not be traced back to them.

After training, the CHWs were free to suggest the system to any care recipients that they interacted with. All CHWs told us they advertised the system to households they visited, and one advertised it at a local hospital. We are aware that asking CHWs to recruit people to submit feedback regarding the CHWs' own work may lead to bias [55], such as CHWs telling people to submit positive feedback. We ideated on ways to try and mitigate such bias. For example, we focused on training CHWs to (1) teach beneficiaries to use the USSD system on their own, and (2) emphasize that beneficiaries could submit feedback at any time (such as after the CHW departed).

To further encourage participation, we also sent SMS reminders to care recipients. To do this, we collected the phone numbers of care recipients that CHWs visited and asked the CHWs to inform these care recipients that they would, after the home visit, receive an SMS reminder encouraging them to provide feedback about their CHW visit. An example of an SMS reminder is, *"Hello, you were recently visited by a CHW. Please dial *384*888# to provide feedback about the visit. This service is free. Thank you"*. We collected roughly 20 new phone numbers every two weeks, sending a total of 80 SMS reminders during the deployment to care recipients who had received a CHW visit.

Focus Groups. Two weeks into the deployment, we conducted six focus groups with 42 participants: 24 care recipients, five CHWs, three supervisors, and 10 hospital decision makers, to collect qualitative feedback about the deployment. Two CHWs, all supervisors and one decision maker had also taken part in the design phase; however, the remaining participants were new to the study. We recruited participants through supervisors who reached out to their CHWs and CHWs in turn reached out to their care recipients. In addition, one of the research team

Category	**Total**: 42 Participants **Beneficiaries**: 24, **CHWs**: 5; **Supervisors**: 3, **Decision Makers**: 10
Age	**Beneficiaries**: Min: 20, Max: 49, Avg: 31; **CHW**: Min: 33, Max: 46, Avg: 40; **Supervisors**: Min: 30, Max: 40, Avg: 35; **Decision Makers**: Min: 30, Max: 60, Avg: 40;
Gender	**Beneficiaries**: Female: 21, Male: 3; **CHW**: Female: 3, Male: 2; **Supervisors**: Male: 3; **Decision Makers**: Female: 6, Male: 4;
Education	**Beneficiaries**: form two - diploma; **CHW**: form two - college; **Supervisors**: diploma - masters; **Decision Makers**: diploma - medical degree;
Phone	**Beneficiaries**: basic phone: 17, smartphone: 6, shared: 2; **Other participants**: smartphones;
Beneficiary Occupation	farmer: 10, trader: 4, house wife: 5, teacher: 2; business owner: 1, tailor: 1;

Table 3.2: Demographic characteristics for focus group participants.

members who resides in Siaya invited decision makers to partake in our study. Then all interested participants who came forward participated in different focus groups. Focus groups lasted for about an hour for care recipients and 30 minutes for other stakeholders and questions were tailored to each focus group. For care recipients, we aimed to understand their experiences and challenges of using the USSD system; for CHWs, we focused on how they perceived the feedback and their experience informing care recipients about the system; and for decision makers, we inquired about the role and impact of feedback on community health programs. All participants except decision makers were compensated $5.00 to cover their transportation to the focus group. Table 4.1 shows the demographic details of participants. All care recipients we interacted with were literate and had their own devices. Only two people said they shared their phones. Focus groups were conducted in English and Dholuo by three researchers. In total, we had five hours of focus group discussions.

Data Collection and Analysis. Our qualitative data consisted of focus group discussions and 22 pages of notes that we took during our fieldwork. We audio-recorded and transcribed our focus groups. We then performed thematic analysis [234] on the transcripts and field notes, beginning with a close reading of the transcripts and allowing codes to emerge from the data. Multiple passes through the data resulted in 39 codes (e.g., *bring about change* and *concerns about negative feedback*). We clustered related codes into high-level themes (e.g., *promoting equity* and *sustainability*) and organized them in a codebook. After multiple discussions and iteratively refining the codes, we arrived at a final set of themes that comprehensively represents the data.

Our quantitative data consisted of system usage logs recording when and how often users dialed into the system. For each user we recorded the choices selected, any errors made (and types of errors), time spent on each screen, languages chosen, types of feedback submitted, and timestamps of all events. All data logged during training sessions were removed from the analysis. We also recorded when SMS reminders were sent and when users responded, if they did.

3.8 Findings from Field Deployment

We now discuss our deployment results as they relate to two key themes: (1) how the system was used by care recipients, including the kinds of feedback submitted, the effects of training, and the impact of SMS reminders; and (2) the socio-technical factors impacting our deployment.

3.8.1 System Usage

Overview. During the deployment, we observed varying levels of engagement with the *Hospital Line* and the *Household Line*. In total, we recorded 495 sessions and 2602 hops. A session is defined as a period from when a user initiated a code to when it ended, while a hop is a USSD request that shows the user a single question. Over the deployment, 168 unique phone numbers dialed into the system (79 into the *Hospital Line*, 124 into the *Household Line*, and 35 into both lines). Figure 3.3 shows the number of times users dialed into the system. Most users (n=64) dialed in only once. The median number of dial-ins was two times, and the maximum was 17 times.

It is important to note that users who dialed into the system did not necessarily complete the entire survey. Indeed, Figure 3.3, which shows the number of hops completed by each user who initiated a session, reveals that the number of submissions decreases as the number of hops increases. The tendency of completion rates to decrease as questionnaire length increases is well-documented in the literature (e.g., [80]). In our case, the dropoff may be due to several factors, including USSD timeout errors or people running out of time, losing interest, or dialing in out of curiosity. Nevertheless, a benefit of USSD is that the system records data from partially completed surveys (any completed hops). As a result, our total data set is larger for some questions than others.

The decrease in question completion rates also suggests it is advantageous to place the most important questions first. However, in our design, we asked simple, demographic questions first (e.g. gender, age) so that people would find the system easy to use and feel encouraged. This meant that the important feedback rating question was positioned towards the end of the survey (see Figure 3.2) and completed by fewer people. We plan to correct this in future

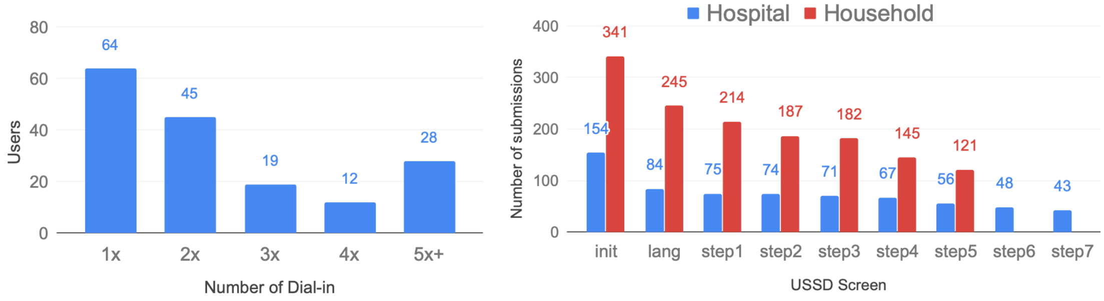


Figure 3.3: Left: Frequency of individual user dial-in (median = 2). Right: Submissions for each USSD screen. The user starts from initializing the code (init) to selecting a language (lang) and then going through seven or five steps for the hospital and household lines respectively.

USSD Line	Sessions	Phones	Women	Men	Avg Age	Min Age	Max Age
Hospital	154	79	92	34	30.8	19	70
CHW	341	124	179	59	33.3	18	70
Total	495	168*	271	93	32.1	18	70

Table 3.3: Details of USSD submissions: *"Sessions"* refer to each unique time that a user dials in; *"Phones"* refers to the unique phone numbers that dialed in (some phone numbers (n=35) submitted to both the Hospital and Household lines). *"Women"* and *"Men"* show gender indicated on submissions; Age is in years.

design iterations.

Error rate. Our system was designed to minimize erroneous user entries. Whenever a user entered an invalid input, they received a message that gave them the range of values allowed. For example, when a user is asked *"Your age"* and enters non-numeric text, the system prompts: *"Please enter a number between 18 to 99"*. This prompt persists until the user enters a valid value. Surprisingly, the error rate in the system was negligible: 0.6% (16 out of 2602 hops) from eight phone numbers. This was in part due to our **behavioral design elements** that simplified the interface for users.

All errors were of two kinds: (i) entering text when only numeric responses are allowed; and (ii) submitting numeric data outside the input range allowed. The overall low error rate is in part due to: (i) familiarity with USSD—participants said they were familiar with USSD; and (ii) understandable questions—participants said that the questions were straightforward and easy to understand.

The kinds of feedback received. We received a total of 193 responses to the Net Promote Score (NPS) question: 135 (69.9%) positive, 33 (17.1%) negative, and 25 (13.0%) neutral (see Fig. 3.4). Most feedback (145 submissions) came from households. We computed the NPS [51] for both lines (% positive feedback - % negative feedback). For hospital feedback, 58.3% (n=28) were positive and 25% (n=12) were negative, yielding an NPS score of 33.3. For household feedback, 73.8% (n=107) were positive and 14.5% (n=21) were negative, yielding an NPS score of 59.3. Thus, household feedback was more positive than hospital feedback. There are several possible explanations for this. For example, for household visits, CHWs make the effort to travel to beneficiaries' homes, which is easier and more convenient for beneficiaries than traveling to and waiting in line at the hospital. Thus, a household visit may actually be a more positive experience than a hospital visit. In addition, CHWs often have personal relationships with beneficiaries, which is not necessarily the case for staff at hospitals. Finally, household visits may have been more prone to response bias [55] since CHWs were soliciting feedback about their own services.

Training leads to engagement. Recall that CHWs were trained to guide care recipients on how to submit feedback via the USSD system. Our findings show that CHWs were comfortable training care recipients. On the first day of deployment, our team accompanied a CHW to the laboratory department of a hospital and ob-

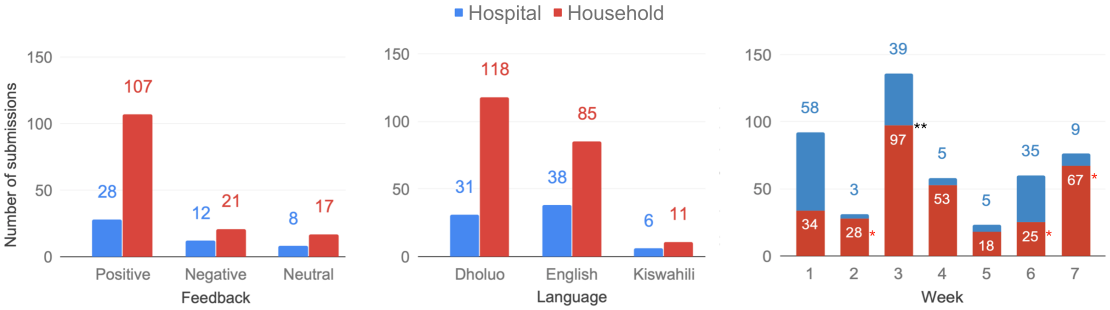


Figure 3.4: (Left) The kinds of feedback submitted; (Middle) Language selection; (Right) Submissions per week. SMS reminders were sent to households in weeks two, six and seven (*). In week three (**), we conducted focus groups, and observed an increase in submissions on both the *Hospital* and *Household* lines.

served from a distance as the CHW spoke to a group of hospital attendees about how they could provide feedback after their hospital visit. Within three hours, 16 users had dialed into the system and made 20 submissions, demonstrating that CHWs may be a good channel through which to reach care recipients. All CHWs said that it only took a few minutes for them to explain to care recipients how to use the system. In addition, they emphasized that care recipients could submit feedback at any time. One CHW shared,

> *"Training care recipients was OK. I did it and it was easy for me and my households. But there was a challenge. They asked me 'Are we going to be given so many questions to answer and questions that we don't know how to answer?'. So I just tried to tell them 'you'll be given questions that you will be able to answer."'* (P27, Female, CHW)

In addition, care recipients shared that after they were informed about the system, they were able to dial into the system by themselves and submit feedback because they were familiar with how to use USSD. However, CHWs also said that, occasionally, they visited care recipients who did not possess mobile devices. In these cases, they did not mention the feedback system. As a result, training only occurred when CHWs felt it was appropriate. One CHW shared,

> *"There are some households where you find that the care recipient is an old mama and she doesn't have a phone so you find that there is no need training her about the USSD tool or her submitting feedback." (P25, Male, CHW)*

SMS reminders renew engagement. We sent SMSs to household care recipients reminding them to submit feedback about recent CHW visits. We sent a total of 80 SMS reminders during the deployment. SMSs were sent on weekdays, between 10am and 4pm, the week after care recipients had been visited by a CHW. Of the 80 SMSs sent, 75 were successfully delivered, while five failed because of network issues or the numbers being out of service. Figure 3.4 shows that, on average, system usage increased after SMS reminders were sent. We received feedback submissions from 20 out of the 75 people who received an SMS reminder (27%). The fastest response was received 3 mins after the SMS was delivered, and the slowest 21 hours after the SMS was delivered (average: 2.5 hours, median: one hour). However, not all SMS reminders triggered engagement for a variety of reasons, such as the SMS not being delivered or care recipients not wanting to submit feedback, perhaps because they did not have time or felt indifferent about the services received.

3.8.2 Socio-Technical Factors Impacting Deployment

Having described how care recipients engaged with the system, we now discuss socio-technical factors that affected our deployment. We uncovered five such factors: (1) trust and accountability, (2) privacy and anonymity, (3) equity (4) leadership buy-in, and (5) feedback granularity.

Trust and Accountability. Our findings showed that feedback from care recipients fostered a new layer of social interactions in the community, helping to promote accountability and build trust in the way health services are delivered. All care recipients and CHWs expressed that the availability of the *Hospital Line* and the *Household Line* served, in combination, as an empowerment platform for care recipients to have their voices heard in ways that were not possible before. For example, participants described how, previously, when they were dissatisfied with services at hospital facilities, they bottled up their complaints because they did not have a way to share their experiences. Participants were now happy to have a way to report feedback. One said,

> *"The majority are happy with the USSD tool and they accept it. They are really happy with it, more so in the hospital. They really appreciated that because they normally face people who have got many characters. For example, some people mistreat care recipients in the hospital laboratory department. Or at another department they normally take long. But if there's something that is being done in the community such that the client reports it and it can be followed up, then at least the people responsible are going to change."* (P28, Female, CHW)

Hospital decision makers explained that they wanted to collect client feedback to gain insights into how people were treated across hospital departments. They thought that although some clients might have poor experiences at hospitals, these could perhaps be attributed to one or a few specific departments. Feedback might help to identify such departments and hold them accountable.

With respect to feedback about CHWs at the household level, many participants (n=13) expressed that health was an important matter that required attention. As such, feedback might be useful in discouraging CHWs from reneging on their responsibilities. One participant said,

> *"[CHWs] have to do a good job because if they don't, they know we shall be submitting feedback about it... Since the work of CHWs are issues that are concerned with health, it is important they do a good job. With good health my life is okay, so I should be there to speak the truth about their work."* (P17, Male, Care recipient)

However, CHWs worried that they might receive unwarranted negative feedback from malicious users in their communities. Some participants (n=7) shared that even though CHWs worked hard, they sometimes encountered difficult community members who did not value their contributions. CHWs worried that such individuals could easily use the system to leave negative reviews. A CHW said, *"It's like we are selling what is maybe going to kill us."* (P25, Male, CHW). We assured CHWs that our research was exploratory and made clear that any feedback received during our study would not affect their jobs. However, these tensions raise open questions for future research.

All CHWs (n=5) also felt that, although positive feedback encourages them, genuine negative feedback could help them know how to improve their work. They described wanting to be able to personally review all the feedback received about them. One CHW shared,

> *"It's now upon me as a CHW to take all the positive and negative criticism and work on them either to scale up my work or to improve on what I have been doing. Even when we were telling our fellow CHWs, we told them that it is not a tool that has been introduced to destroy what we have been doing. It's just to monitor and to make us aware of our weaknesses and to help us improve on those weaknesses. One or two people might use it negatively*

> *against us but not everybody will do that. I hope it will make us improve on what we are doing."* (P25, Male, CHW)

As discussed above, we received much more positive feedback than negative feedback (see Fig 3.4), which suggests that care recipients are perhaps more likely to submit positive feedback.

Privacy and Anonymity. Our analysis revealed that privacy played an important role when providing feedback. All participants said they were confident they could provide feedback without someone else knowing they had done so. One participant shared,

> *"When I am done answering the feedback questions, the [dialog on the] screen disappears—it's already gone and no longer on the phone."* (P21, Male, Care recipient)

When a user interacts with the system, the pop-up dialog on their screen is immediately removed when (i) they complete the last USSD question; (ii) they press the cancel button to exit before the last screen; or (iii) the mobile network automatically makes the screen timeout if the user has taken longer than one minute. Once the dialog has been removed there is no visible trace on the user's device that indicates they dialed into the system. As such, USSD automatically makes users' interactions private compared to systems that use SMS (where users would have to delete each SMS to hide their interactions) or phone calls (where users would have to delete their call logs).

Hospital decision makers and a few CHWs (n=3) inquired if we could tell which care recipients submitted feedback. We explained that we could not because the system further protects user privacy at the backend. All feedback received is stored in a database that does not contain any user names. We do, however, store user phone numbers as a way to distinguish return users from new users. As a result, an administrator who looks at the data cannot infer who submitted feedback unless they have a database that connects phone numbers with care recipient names.

Equity. Our design promotes equity by making the system free to use by anyone who dials in. All our participants had access to a personal or shared phone and Safaricom SIM card. Several participants told us that they deliberately checked that the system was indeed free to use. One said,

> *"It is free and I didn't use any money. I know because after dialing into the system, I checked my airtime balance."* (P3, Female, Care recipient)

If a user dials into the system from a network other than Safaricom, they receive an error message saying that the code they dialed is invalid. A few participants (n=5) asked if it was possible for the system to work on all networks. We explained that since we were in the exploratory phase of our work, we limited the system to Safaricom for financial sustainability (discussed more in Section 3.9).

In addition, participants shared that the opportunity to choose their own language upon dialing into the system made the system more accessible and easier to use. Figure 3.4 shows that Dholuo was the most preferred language, followed by English and then Kiswahili. The low usage of Kiswahili was consistent with participants' feelings during our focus groups, where they described how they were not comfortable speaking or reading Kiswahili because it was not their mother tongue.

Leadership buy-in. Our design approach purposefully involved hospital and community leadership from the start. At the beginning of our project, we reached out to the Siaya County Ministry of Health leadership requesting to test the system at one hospital, approaching clients as they exited the hospital. In response, we were invited to test not only at the exit of the hospital but also in the waiting area inside a hospital department. This perhaps indicates these stakeholders' enthusiasm for obtaining client feedback. We kept these stakeholders updated about subsequent iterations of the system, and were subsequently invited to partner and expand the deployment to ten hospital facilities in Siaya. Although the hospital decision makers were familiar with well-known data collection tools such as ODK (Open Data Kit) [32], they were enthusiastic about using USSD for collecting community feedback because it was accessible by community members with basic mobile phones and, unlike ODK, it did not require extensive technical support or training to deploy. We are now in the process of coordinating with stakeholders to expand the system to multiple facilities.

Granularity of feedback. A key finding in our research is that it is complex to collect granular feedback. Although it is relatively straightforward to collect feedback about how communities view CHWs in general, understanding which communities submitted feedback would require users to identify their community (e.g., village) in their submission. However, as discussed in Section 3.6, we eliminated text-based responses to avoid frequent timeout errors. Alternatively, users could choose from multiple choice options of villages, but there are hundreds of villages and the list would not fit on a single screen. Compounding the challenge of getting granular feedback is accounting for scenarios in which multiple CHWs visit one household. CHWs explained that although each CHW is in charge of a set of households, it was common for a household to receive services from CHWs that work in different health organizations. One CHW explained,

> *"We go across each other's borders. You might find that there's a CHW dealing with a household on a different issue. Maybe there's a different organization that has come and recruited their own CHWs ... Does the feedback collected apply to all CHWs or only the particular ones in our area? Can the care recipient give feedback since the tool is not particular about who specifically visited the recipient?"* (P25, Male, CHW)

This suggests that separating feedback about multiple CHWs that attend to a household will be challenging if there is no unique identifier that links specific CHWs to the feedback submitted.

3.9 Discussion of Field Deployment

Having analyzed how care recipients engaged with the USSD system and the socio-technical factors that impacted our deployment, we now discuss four key themes to address as we move forward: the feasibility of using USSD to collect feedback from communities, how the feedback may be used, how we might scale the system to larger deployments, and challenges impacting sustainability.

Feasibility of USSD. Our analysis suggests that USSD has a strong prospect of becoming an effective mechanism for collecting community health feedback from care recipients and contributes to the limited literature available on USSD-based systems [193]. As a universal communication channel embedded in all mobile devices, it is accessible on any basic feature phone, and access is not dependent on Internet connectivity. Since our participants in Kenya were already familiar with using USSD for mobile carrier services (e.g., checking their airtime balance), it was easy for them to learn how to apply the same approach in providing community feedback. Surprisingly, our data revealed an extremely low error rate (less than 1%), which corroborates our qualitative findings that users did not struggle to use the system. Informing our design process with **behavioral science concepts** played a big role in simplifying the interface for users and making the process of sharing feedback intuitive. This approach is uncommon in the HCI4D literature yet it is beneficial. For example, our design which led to low error rate contrasts prior research that describes high error rates in SMS-based systems in low-resource contexts [195, 188].

Our stakeholder-engaged approach led to a number of key design decisions that aided deployment. For example, making the system free to use promoted equity by enabling anyone with a basic mobile phone and a Safaricom SIM to submit feedback. Providing the system in multiple languages further aided access; we saw usage in all three languages, with Dholuo (the local language) being most popular. These findings support prior work on the importance of local-language settings [57, 192, 193, 260] and extend this literature by showing the benefits of supporting *multiple* languages.

Although USSD is a promising approach for collecting community feedback, it also has limitations and challenges that could impact adoption. Clearly, it requires users to read the questions on the screen and thus may be inappropriate for people who are illiterate. Beyond this, our analysis revealed technical challenges with adopting USSD. For example, each screen was automatically limited to an interaction timeout of one minute, and the maximum length of a single screen was 154 characters. We accounted for these constraints through careful design, although coming up with a design in which each screen was less than

154 characters regardless of language was challenging.

Actionability of feedback. This study is part of a larger program of work on precision performance management in community health, which aims to establish new, data-driven approaches to supervision and health system leadership. While our study explored a system for collecting beneficiary feedback, in our interviews with health system leaders we discussed possible channels for making this feedback actionable within existing infrastructure. We now share ideas about how such a system might be useful and what actions may be taken based on the feedback.

Prior studies have shown that giving CHWs access to feedback about their work can lead to performance improvements [59, 262]. For example, showing personalized performance feedback to CHWs via a dashboard during face-to-face meetings with their supervisors led to improvements in CHW performance in Mali [262]. However, the feedback shown to CHWs did not include subjective feedback from care recipients. Thus, a concrete use case would be to add to the dashboard aggregate feedback (e.g., ratings) submitted about the CHW. Supervisors and CHWs could then discuss this feedback and any actions to take during their face-to-face meeting. In addition, we worked within a CHW program that already presents CHWs with targets (e.g. targets for facility-based births among women enrolled in antenatal care) built using the open source Community Health Toolkit [42]. Thus, another use case would be to create new CHW targets for care recipient feedback (e.g., number of feedback reports the CHW receives, target quality ratings, etc.) CHWs could then view aggregate feedback submitted by their care recipients and use it to improve their work (e.g., by making more household visits or improving the quality of the visits that they perform).

To make the feedback actionable for community health program leadership, we note that the decision makers we worked with already use the open source health management information system DHIS2 [62]. Using DHIS2, decision makers routinely view a range of dashboards designed for high-level health system management. We anticipate building a care recipient feedback view within these dashboards that would allow decision makers to see aggregate feedback about CHWs and health facilities. By including questions that ask care recipients who submit negative feedback to choose from a list of common reasons (such as *"waiting time is too long"*, *"medications are too expensive"*, *"facility staff are rude"*, etc.), the system could provide decision makers with possible next steps to take based on the feedback (such as more staff training, increasing stocks of pharmaceutical supplies, subsidizing the cost of medications, etc.). Further, decision makers could update the survey questions periodically to seek feedback on specific topics of interest, such as asking *"were you treated with respect"*, *"did you face discrimination"* and so on. New questions would still need to be carefully designed and pilot tested prior to large-scale deployment.

For each of these scenarios, it will be important to carefully consider privacy, who has access to the feedback, and how it might impact CHW and hospital staff employment (e.g., if they could lose their job). This will be particularly important as we explore opportunities to link feedback received to individual CHWs or health facilities. One technique for linking feedback in this way would be to assign unique ID numbers to CHWs and facilities, which beneficiaries would enter into the USSD system. This approach is similar to how mPesa customers submit a unique number into mPesa in order for mPesa-accepting business points to be identified [104]. The IDs for health facilities could be displayed on posters in each facility and distributed to care recipients by hospital staff; while CHWs could share their IDs with care recipients during household visits. Implementing identifiers in this way would open up new opportunities for synthesizing insights across different data sources. For example, the IDs could be used to connect aggregate feedback data with other performance information such as the speed with which CHWs reach sick children. Our next study plans to explore how to assign unique identifiers to CHWs and health facilities that are shared with care recipients as part of a broader aim to incorporate beneficiary feedback into existing infrastructure for precision performance management.

Scalability. We conducted a small-scale deployment, training only five CHWs who trained community members. We did this because we wanted to understand the impact and consequences of the system before proceeding to a large-scale deployment. For instance, CHWs were worried that beneficiaries might maliciously submit negative feedback that would impact their employment, which our small-scale deployment suggests is perhaps not a major concern (at least for now).

Our findings show that sending SMS reminders often prompted care recipients to engage with the system and submit feedback. This finding corroborates prior research that SMS reminders trigger engagement in community health programs [58, 192]. Although manually sending SMS reminders to a few households per week was manageable in our small-scale deployment, this approach would not be suitable at scale. In addition, we sent reminders at the beginning of the week after care recipients received a CHW visit because we did not want to send reminders to care recipients if they had not yet been visited. However, this led to scenarios where care recipients received reminders to submit feedback several days after a household visit.

As we consider scaling the system, we intend to create an automated SMS management system that is integrated with the existing information system that CHWs already use to collect data about clients they visit (in our case, Medic Mobile [169]). This would enable SMS reminders to be automatically sent to care recipients immediately after CHWs complete their household visits and submit the data to Medic Mobile. This would ensure that the visit is still fresh in the care recipient's mind when they provide feedback. A few days after prompting for feedback, the SMS system could check the USSD system to see if care recipients submitted any response and if not, send an additional reminder message. We elected not to integrate an automated SMS reminder system in our initial deployment because we were still unsure about the feasibility of the USSD system.

Finally, we conducted a small-scale deployment with limited participants in one county in Kenya. More research is needed to evaluate the system at a larger scale and in different contexts.

Sustainability. A key part of scaling our USSD system is sustaining the deployment over a long period of time. Sustainability remains a longstanding problem for researchers and practitioners who deploy projects in low-resource settings [256, 192, 250, 260]. Our analysis showed that setting up and maintaining a USSD system in Kenya is significantly cheaper than setting up an IVR system (by a factor of 10) but slightly more expensive than setting up an SMS system (by a factor of two). We estimate the cost of setting up two USSD lines to run for a year to be $1700. This would account for 20,000 free feedback submissions to the USSD system. Based on these figures and in collaboration with our research partner, Medic Mobile, we plan to deploy the system at a larger scale and in a broader set of communities. In this deployment and beyond, Medic Mobile plans to support the system by integrating it into their existing, widely-used open source software platforms [42, 169].

Beyond financial sustainability, effort will be needed to consistently publicize the system within communities and train users on how to submit feedback. Taylor et al. [239] emphasize that usage issues are a primary challenge in the handover of community health technologies because *"a technology can be given to the community but might not be used"*. As a result, sensitizing communities about the need to provide feedback is important for sustainability. Creating awareness could happen at hospital facilities through ongoing health education activities where community members gather to learn about how to improve their health. Catchy posters could also be put up in strategic places, such as waiting areas and exit gates of hospital facilities. At the household level, we worked with only five CHWs who reached households in Siaya. Effort would be needed to train more CHWs, who in turn could train community members. Communities could also learn about the system through community events and meetings that take place in many villages, word of mouth at household levels, and radio stations. However, these combined efforts to sensitize communities on the need to provide feedback on their health programs would incur additional expense.

CHAPTER 4

CASE STUDY 2: DESIGNING TECHNOLOGIES FOR HOME HEALTH AIDES CARING FOR HEART FAILURE PATIENTS IN NEW YORK

4.1 Introduction

This chapter begins with a qualitative research that identified the diverse groups of stakeholders in a heart failure ecosystem and concludes with results of testing a digital provocation application with these stakeholder groups. Home health aides (HHAs) are an important group of frontline health workers and one of the fastest-growing sectors of the US workforce and healthcare industry [232, 196]. There are currently two million HHAs in the U.S. [33] and, between 2014 and 2024, HHAs will add more jobs to the US economy than any other occupation [107, 33]. Largely employed by agencies receiving public funding (Medicare and Medicaid), HHAs care for 48 million Americans and account for 74 billion dollars of healthcare spending per year [196]. Prior work shows that HHAs work in patients' homes, helping them to manage a wide range of chronic diseases and navigate the healthcare system [8, 77, 52]. In particular, many HHAs provide long-term assistance and post-hospitalization care for adults with heart failure [119, 118, 158].

Caring for patients with heart failure is a pressing concern. In 2013, heart failure became the leading cause of hospitalization in the US and Europe [2]. In the US, there are one million heart failure-related hospitalizations per year and 25% of patients are readmitted within 30 days [115, 135, 38]. Frequent re-hospitalization contributes to patient morbidity and mortality, and also financially impacts hospitals since, beyond the cost of the hospital visits themselves, under value-based healthcare reform in the US, hospitals now incur additional financial penalties for 30-day patient readmissions [83].

HHAs may represent an important but untapped opportunity to improve patient outcomes, such as reducing re-hospitalizations in heart failure. Unlike physicians or visiting nurses, HHAs are with patients and in their homes on a near-daily basis, up to 24 hours a day, which gives them a unique vantage point from which to observe, assist, and advise. They are often central in the lives of their patients, assisting with meal preparation, medication compliance, physical activity, symptom management, attending medical appointments, and more [100, 233, 8].

Yet, to date, they have not been the focus of research in heart failure or interventions to improve patient outcomes [231]. They do not receive educational resources to help them learn about heart failure [230, 207], and struggle to reach their supervising nurses and patients' doctors when patients are symptomatic and they need clinical help [230]. Beyond the difficulties they face in providing care to heart failure patients, HHAs are themselves a vulnerable and marginalized workforce. Although one of the fastest growing sectors of the healthcare industry, HHAs, who are often women and minorities, endure erratic employment, low wages, discrimination on the job, and high levels of burnout [118, 119, 21]. Indeed, prior work has characterized the home care workforce as *"invisible"*, *"ubiquitous"*, and *"continually undervalued"* [233].

We hypothesize that digital technologies have the potential to address some of the educational, communication, and equity challenges that HHAs face, thereby improving healthcare delivery for heart failure patients as well as the sustainability of the US healthcare system more broadly. However, little is currently known about if or how technology currently impacts HHAs' work practices as they care for heart failure patients in the home. To this end, we contribute a multi-stakeholder qualitative analysis with 38 participants that examines the role of technology in HHAs’ care of heart failure patients in New York City (NYC). Through 17 semi-structured interviews and seven focus groups, we sought to understand how technology is used, why it is used in the way it is, particularly in the context of existing organizational rules and infrastructures, and how stakeholders feel technology could better serve them. Although HHAs were our target users, we gathered the views of multiple groups of stakeholders to provide a rich and balanced perspective of how technology is integrated into day-to-day activities.

Our findings show that when a HHA begins working with a heart failure patient, they become part of a complex set of socio-technical systems, digital and non-digital, that enable them to deliver care. The HHA begins by working with a nurse to derive a set of tasks that should be performed for the patient and that are recorded on a paper-based care plan that remains in the patient’s home. Any tasks that HHAs accomplish while working with a patient are then reported at the end of their shift by calling into a telephonic punch-code task reporting system using the patient’s in-home telephone. These paper care plans and telephonic punch code systems constitute the entirety of the tools purpose-built to support HHAs’ work. However, to fill in the gaps between these tools and the complex needs of modern home care, HHAs in our study reported using their personal mobile devices for a range of activities (despite patient privacy concerns) including incident reporting, calling emergency services, searching the Internet for information, and recording and storing patient information necessary for care.

We discuss the ways in which these existing socio-technical systems challenge or disadvantage HHAs in their work by systemically deprioritizing their needs and perspectives. For example, we show existing information systems focus primarily on monitoring HHAs rather than providing tools to support their work. Compounding this is a lack of integration between home care and medical teams, as well as a lack of investment in sustainable information architectures. In light of these challenges, our participants shared their ideas for how technology could be used to better serve HHAs' needs, such as revising the task recording system to become more flexible and dynamic, enabling better communication between care teams, and enhancing agency management systems.

Building on these insights, we constructed a design provocation: a non-functional prototype of a tablet application that appeared to center aides' needs and provide them with resources for educational content, improved communication with their supervisors, and improved ability to record their daily tasks. We then used this provocation in a field study with 16 participants (aides, nurses, and aide coordinators) to provoke an exploration of currently existing and possible future design spaces around aides' work.

Findings from testing the provocation surfaced key insights into the perspectives of aides, nurses, and coordinators as they interacted with our provocation. Aides perceived the provocation as a way to gain more control over their work, but they clashed with nurse and coordinator participants in their desire for better communication. Participants were similarly mismatched in their opinions of providing aides with educational resources. Across stakeholder groups, participants also perceived the provocation to be a trigger for conflict in clients' homes. Finally, participants expressed contrasting levels of concern about whether and how aides might respond to the added workload and learning curve that the provocation might represent.

We conclude by demonstrating how our work provides a concrete enactment of Drombowski et al.'s [68] framework of social justice-oriented interaction design, and discuss how our work extends three key strategies within this framework: 1) the ways in which our provocation explored designing for *transformation* as a way to see how near-future political shifts might change marginalized workers' experiences; 2) the need for designers who seek to *enable* marginalized workers to weigh the heightened expectations created by such projects; and 3) the difficulty of designing for *reciprocity, distribution, and accountability between stakeholders* within complex multi-stakeholder environments like the one in which we worked.

4.2 Related Work

4.2.1 Technology and Heart Failure Management

Prior work explores the use of technology for heart failure management, including tools for monitoring physiological symptoms [143, 161, 212]. In the medical literature, researchers have explored how technology broadly supports home care support. Cipriano et al. [41] described recommendations for designing technologies that support collecting, analyzing and sharing information with providers, patients and caregivers. Matthew-Maich et al. [164] performed a scoping review of mobile technologies for managing chronic conditions, including heart failure, and pointed out the limited use of technologies in homes. Morey et al. [174] looked at the challenges of common mHealth apps designed to manage congestive heart failure and highlighted design issues that limit usability.

However, this prior research on heart failure management has focused on tools that facilitate interactions with doctors, nurses, and patients – not paid HHAs. A systematic review found only six out of 7,032 studies focused on

HHAs, revealing that this workforce has largely been neglected in research and interventions around heart failure [231]. This work also showed that HHAs feel overworked and undervalued, experience a myriad of challenges caring for heart failure patients, and find heart failure management to be frightening and unpredictable because it involves life-and-death situations [231].

4.2.2 Informal Caregiving for Chronic Diseases

Beyond heart failure, a substantial amount of HCI research has examined the management of chronic diseases such as diabetes [60, 180, 74], dementia [263, 147], cancer [75, 114], and more. Interest in chronic disease management has grown with increasing attentiveness to post-hospitalization and long-term assistance provided through informal caregiving (i.e., caregiving by patients' family members). For example, Kaziunas et al. [125] studied the interconnections between information and emotion work performed by parents as caregivers. Schorch et al. [213] gathered data from observations, interviews and cultural probes to gain a better understanding of the practices of relatives as caregivers. In their qualitative research on family caregivers and patients with multiple chronic diseases, Lim et al. [150] combined interviews, photo elicitation, and home tours to explore the design of socio-technical tools that improve patient-provider communication. More recently, HCI researchers have begun to explore how technologies could play a role in these efforts. Yamashita et al. [268, 267] explored the use of tracking technologies to assist family caregivers caring for depressed patients. Nunes et al. [178] performed a systematic review to understand how patients and carers use self-care technologies.

However, most HCI research on caregiving has focused only on informal caregiving by family members, with little attention on formal, paid caregiving. Our paper expands this body of work on caregiving by providing an in-depth analysis of how formal, paid HHAs use technology to care for adults with heart failure.

4.2.3 Health Services by Frontline Workers

The global impact of frontline workers delivering health services to hard-to-reach communities has been well studied across diverse contexts and communities [14, 57, 37, 192]. Many health programs focused on improving the health of underserved populations increasingly rely on frontline workers, particularly in low- and middle-income countries, because these paraprofessionals can positively cause behavior change and reduce neonatal mortality rates [14, 141]. As such, a number of tools have been designed to support frontline workers in community health programs to achieve their health objectives. For example, researchers have designed automated SMS systems to boost engagement in community health programs [192, 58]. DeRenzi et al. [57] created a personalized feedback system for frontline workers in India that led to more than 20% increase in the average number of client visits performed by each worker. In addition, researchers have explored diverse techniques for collecting useful feedback from users when exploring the design space of tools for frontline workers and community health programs. These techniques include adopting dramatic story lines [37], role playing with skits [173], and exposing people to their neighbors' critical feedback to persuade them to act similarly [255].

However, this prior work has primarily focused on community health programs in low- and middle-income countries, with little HCI research examining community health programs in so-called "developed" or resource-rich regions (such as HHAs in New York City). Outside of HCI, a growing number of studies in the medical literature have showed that interventions by frontline health workers in resource-rich locations (e.g. the United States) can improve the health outcomes of underserved populations [113, 257, 154, 124].

Our paper extends prior research with an empirical study of how paid, formal caregivers (HHAs), who are themselves a marginalized and vulnerable workforce, use technology in their day-to-day work caring for adults with heart failure. Our study yields rich insights and design recommendations for researchers and system designers interested in creating technical tools to aid the management of heart failure and potentially other chronic diseases.

4.2.4 Justice-Oriented Design Sensibilities

Our work draws on growing research in HCI and interaction design seeking to establish theoretical foundations of how designers can work to improve the lives of underserved, marginalized, or oppressed communities in the face of large-scale systemic inequities. Much of this work (e.g., [15, 109]) has outlined sets of design sensibilities that stand in contrast to the dominant paradigm of HCI as a practice of engineering technology solutions in response to commercial or user requirements. The expansion of the design space beyond individuals' immediate needs to collective social problems requires a corresponding expansion of designers' approaches. We locate our work on this frontier.

Principally, we are compelled by the framework put forth in Dombrowski et al [68], outlining three personal commitments and six strategies through which designers might practice a social justice orientation. Such an orientation, the authors posit, helps designers seeking to address large-scale social challenges to ground their work in a landscape shifting from a focus on technological possibility as the prime directive to a focus on understanding how to design ethically, responsibly, and with accountability. Projects developed in this manner start with designers' personal commitments to 1) the *conflict* inherent to a polyvocal design process; 2) the *reflexivity* required on the part of the designer to acknowledge their own biases and how they might color the work; and 3) engaging with the *ethics and politics* of creating change, specifically by taking a personal stance in one direction or another.

With these commitments established, designers proceed along one or more of six strategies. Designing for *transformation* shifts designers' orientation away from addressing immediate, individual needs and towards large-scale, long-term approaches to systemic change. Designing for *recognition* centers on identifying unjust phenomena as a precursor to addressing them. Designing for *reciprocity* aims to create conditions in which relationships within an inequitable system can change to become more equitable for all stakeholders. Relatedly, designing for *distribution* seeks a more even sharing of social systems' benefits and burdens among stakeholders, and designing for *accountability* foregrounds the creation of avenues along which those who oppress and who benefit from others' oppression can be held responsible. Finally, designing for *enablement* aims to foster human capacity and self-determination.

This framework draws on related threads within HCI complicating the paradigm of user-centered design with feminist [16, 15], postcolonial [109], and queer lenses [148]. We note it also engages with questions of self-determination central to ongoing work in the HCI subfield of accessibility, notably Ladner's [142] concept of designing for *empowerment*. Projects with this tenor, Ladner argues, ensure users are active participants or even leaders in every stage of the design process, from defining the problem through to building and evaluating the resultant systems. Ladner posits that HCI projects purporting to empower a marginalized user group must first show that users have the technical ability to lead the work, and then ensure the users have sufficient pre-existing self-determination to carry forward into the design process.

Our work provides an empirically-grounded study of how these theoretical frameworks play out in practice. Specifically, we discuss extensions of Dombrowski et al.'s [68] framework based on learnings from our efforts to conduct social justice-oriented technology design to improve equity for home health aides who work with clients suffering from heart failure.

4.2.5 Enactments of Justice-Oriented Design

Looking to the problems within our research context, we were inspired by work within the HCI and HCI4D communities examining how the tenets of social justice-oriented design might be applied to improving the lives of marginalized workers.

Extensive ethnographic work has built a nuanced picture of worker marginalization within ride-sharing apps in the U.S. [84, 147] and in the Global South [139, 199], pointing out how the technology-mediated gig economy affects workers by creating perceptions of behavior control and expectations of emotional labor [162, 203]. Outside of the gig economy context, scholars have also examined how low-wage workers use technology to address wage theft [67]. Researchers have also adopted an activist stance and deployed technology interventions that seek to directly address these inequalities. For example, Irani and Silberman [110] showed how Turkopticon, a browser extension allowing Amazon Mechanical Turk workers to review their employers, enabled the invisible human labor force behind Amazon's crowdsourcing platform to engage in mutual aid.

Our work joins this growing body of literature extending the theoretical frame-

Category	Count	Gender	Age Range	Responsibilities
All participants	38	M(4), W(32), UR(2)	30-90	—
Agency leaders	10	M(2), W(8)	40-60	Oversee and work to improve care at the agency level
Care coordinators	1	M(1)	40-50	Supervise and manage HHAs' work on a daily basis
Family caregivers	2	UR (2)	UR	Support HHAs in caring for their loved ones
Home health aide (HHA)	12	W(12)	UR	Provide home care to patients with heart failure disease
Patients	4	W(4)	70-90	Adults with heart failure provided with an HHA
Physicians	2	M(1), W(1)	30-50	Care for patients prior to discharging them to home care
Nurses	6	W(6)	40-60	Occasionally visit patients' homes to perform hands-on care
Social workers	1	W(1)	30-40	Work to ensure patients have a smooth home care experience

Table 4.1: Participant demographic characteristics and responsibilities (M: man, W: woman, UR: unreported).

works of design justice through direct application with a group of marginalized frontline workers. The majority of this prior work has focused on settings like gig economies, algorithm-mediated work, and the provision of healthcare in low-resource settings. To this, we contribute a study of frontline healthcare workers in the U.S., a less-studied site for worker marginalization and technology design to combat it.

4.3 Part I: Qualitative Fieldwork on Home Health Aides in a Heart Failure Ecosystem

Our research took place in New York City (NYC) over a six-month period in 2018. We worked with 38 participants from 8 stakeholder groups (see Table 4.1). Before beginning our research, we received IRB approval for all study procedures.

Recruitment and Participants

We recruited participants through snowball sampling and direct outreach to home care organizations. Patients and family caregivers were recruited from internal medicine and geriatric practices in NYC. Participants were interviewed either in person or by telephone. All participants provided verbal or written consent to record the interview with the assurance of strict anonymity. HHAs, patients, and nurses received compensation in the form of a $25 gift card. As Table 4.1 shows, participants ranged from 30 to 90 years old.

Qualitative Methods

We conducted seven focus groups with 21 participants in groups of two to eight people including groups of nurses, HHAs, and agency leaders. We also conducted 17 semi-structured interviews with a nurse, a social worker, a care coordinator, several physicians, patients, agency leaders, and family caregivers. All focus groups and interviews were conducted in English and lasted for 30-60 minutes. The focus groups and interviews were structured around a set of high-level topics and tailored to each participant group. We asked participants about the routines of HHAs during home visits, how patient data was collected, how a patient's health status was shared with health practitioners, and the challenges participants encountered around technology. In total, our dataset consisted of about five hours of focus groups and 11 hours of interviews.

Data Analysis

We audio-recorded and transcribed our interviews and focus groups. We analyzed the data thematically [234], beginning with a close reading of the transcripts and allowing codes to emerge from the data. Multiple passes through the data resulted in 37 distinct codes (e.g., *frequency of communication*, *desired feature*, and *challenges with data*). We clustered related codes into high-level themes (e.g., *desired usage*, *privacy*, and *lack of team integration*) and organized them in a codebook. After multiple discussions and iteratively refining the codes and themes, the research team arrived at a final set of themes that comprehensively represented the data.

4.4 Findings from Qualitative Fieldwork

We begin by describing how technology is currently used in HHAs' care of adults with heart failure. We then highlight specific mechanisms through which the current ecosystem systematically deprioritizes HHAs' needs and perspectives. Finally, we detail participants' suggestions for how their ecosystems could be improved.

4.4.1 The Technology Ecosystem in Home Care

Our research shows that from the moment a HHA begins working with a patient, they are tied into an ecosystem of technologies, digital and non-digital, that enable them to deliver care. Complex and uncoordinated, this ecosystem centers around four core tools: 1) the paper care plan; 2) the telephonic punch-code task reporting system; 3) the HHA's personal mobile device; and 4) the call chain for incident reporting. While our participants described all of these tools as central and commonplace in HHAs' work, our analysis shows that when the first two tools fail to meet HHAs' needs, HHAs turn to the last two tools.

Paper care plans for understanding tasks

Our analysis revealed that HHAs begin their work by referring to a patient's care plan, a guide for how the HHA should assist with the patient's personal and medical care. Written by a visiting nurse early in a patient's episode of home care, the care plan contains such details as emergency phone numbers and a list of tasks the HHA should perform for the patient each shift. While care plans can be updated upon a follow-up visit from a nurse, they can also remain the same for long periods of time.

We heard how this critical document is typically provided on paper, and displayed prominently in the patient's home (i.e. on the refrigerator) so that multiple aides can have access to it. Because they intentionally omit information pertaining to a patient's diagnosis and medical history (due to privacy laws we discuss in a later section), care plans were cited by many stakeholders in our study as a significant source of HHAs' frustration and confusion. Both nurse and HHA participants said care plans were not individualized or specific enough to help aides understand when a given task was disease-related. One nurse participant was also frustrated by the convention of fitting care plans onto one page: *"You can't even read the print because it's so small"* (P1, Female, Nurse).

Telephonic punch-codes for recording tasks

Our participants shared that once a HHA has referred to a patient's care plan, they settle into a daily rhythm of arriving at the patient's home, performing the outlined tasks, and leaving at the end of the shift. Throughout this routine, the HHA's work is tracked and mediated through the mandatory use of a commercially available telephonic punch-code system.

We learned that, at the start of each shift, agencies require HHAs to use their patient's home phone to call into a provided telephone number and follow a series of automated voice prompts to *"clock in"*. In the event that a HHA calls in from their own personal phone, agency coordinators are to follow up with a call to the patient's phone to confirm whether the HHA is on-site. These procedures, participants explained, are to verify that the HHA is in fact with the patient at the time of clock-in.

Participants explained that at the end of each shift, HHAs are required to call into the system again to *"clock out"*, again using their patient's phone. During these calls, HHAs are to report all the activities they performed during their shift by entering a series of numeric codes corresponding to specific actions. A mapping of actions to codes is provided on the patient's care plan. For example, if a HHA prepared a meal for a patient, this task is looked up on the care plan, and the corresponding number, *58*, is punched in response to an automated prompt. Agencies use these records to confirm which tasks HHAs performed during their shift, to understand patients' conditions, and to bill a patient's payer accordingly.

Personal mobile devices for filling in the gaps

The combination of paper care plans and telephonic punch codes constitutes the entirety of the tools purpose-built to support HHAs' care work. However, we discovered how, to fill in the gaps between these tools and the complex needs of modern home care, HHAs in our study use their personal mobile devices for a range of activities. These activities fall into four buckets: (1) reporting incidents and general communication with agency coordinators; (2) calling emergency services; (3) searching the Internet for more information; and (4) recording and storing patient information necessary for care.

First and foremost, participants described HHAs using their personal devices to communicate with their agency coordinators via phone call, photo or video message, or text. It is especially important for HHAs to immediately contact coordinators to **report incidents** like patient falls, sudden weight gain, and more. Upon receipt of an incident report, coordinators are to assess the situation, advise the HHA, and reach out to additional resources if necessary.

Phone cameras in particular were described as a valuable tool for communicating around tasks that were difficult to describe in words like open wound care and cleaning. These tools were especially useful in such contexts:

> *"We had another client that actually got stuck in her bed and they kept trying... The aide kept trying to explain to us how she was stuck in the bed and I was like I don't get it. This lady, how did she get stuck? So they sent over a picture with the name of the bed. We got her another replacement and she did not go to the ER. So it was pretty powerful."* (P1, Female, Nurse)

In cases that might constitute emergencies rather than incidents, HHAs also use their personal phones to **directly call 911**. This impulse is especially strong for heart failure patients, participants described, because the unpredictable nature of the disease means HHAs are sometimes afraid they will be held responsible if a patient is to quickly deteriorate. One participant explained,

> *"The party line was always call 911. I think that call 911 came out of a place of fear that if we don't call 911, what is the impact of doing that?"* (P1, Female, Nurse)

Outside of communication, participants also described HHAs using their smartphones and other personal mobile devices to **conduct general Internet searches** for information relevant to their day-to-day. One nurse called Google search *"my best friend"*, and said she used it to look up more information on health conditions, map directions to patients' homes, and more.

Finally, our data shows HHAs use their personal devices to **record and store patient information**, often in violation of agency policy. HHAs report storing lists of patient medications on their phones for easy access at the pharmacy or at a patient's doctors appointment. One participant mentioned a prior incident in which she had suspected domestic violence at a client site and used her personal device to capture photographic evidence of bruises on her patient's face, in case they became relevant later on. In all cases, participants acknowledged privacy was a concern, but did not report using additional security measures on their personal devices to account for the sensitivity of the information.

Call chain for escalating reports

As discussed above, reporting incidents to agency coordinators is a common practice for HHAs in the field. Our data shows the initial call placed from a HHAs' mobile phone kickstarts a long and complicated call chain process intended to escalate incidents to the appropriate clinical authorities.

We learned that HHAs were to share incident reports by immediately calling their agency coordinators. Then, if needed, the agency was to reach out to an on-call agency nurse. If the on-call nurse needed another opinion, he or she would reach out to the agency nurse who had put together the patient's initial care plan, or attempt to reach either the patient's primary medical doctor or the physician who had overseen their discharge. Ultimately, someone from the patient's clinical team would follow up with either a phone call or a visit to the patient's home to triage. We note that this entire process could take anywhere from hours to days.

4.4.2 Barriers in the Workflow of Patient Care

Having established a clear picture of the tools HHAs use in their day-to-day, we move to understanding the ways in which these technology ecosystems challenge or disadvantage aides in their work. Our analysis shows these ecosystems are currently systemically deprioritizing HHAs' needs and perspectives.

From participants' frustrations, we synthesize three key themes that explain the ways in which HHAs are disadvantaged by their tools: 1) agencies' commitments to information systems that serve as means for monitoring HHAs rather than tools to support their work; 2) the lack of integration between home care and clinical care teams; and 3) a lack of investment in sustainable information architectures. We describe each of these in turn.

Tools for monitoring, not support

Our participants described numerous undue burdens placed on HHAs by their current tools, principally the telephonic punch-code system used throughout the industry to record tasks. These frustrations, participants asserted, stemmed from the fact that the systems were built to monitor HHAs while on the job, not to support them in patient care.

Participants reported frustrations around the *utility* of the telephonic punch-code system, and its *suitability for managing complex cases*, like patients with heart failure. We were surprised to learn the system did not allow task records to be updated if entered incorrectly. As participants explained, since a patient's care plan was not liable to change very much over time, after a few days with a new patient a HHA would begin to enter the same series of numeric codes day in and day out. As the task of recording tasks itself became a matter of rote, the HHA would enter codes rapidly and sometimes make mistakes – but in such scenarios, they would continue on to the next task to be reported, because the only way to edit an erroneous entry would be to end the call, redial, and start the process all over again.

Even when they were able to correctly enter tasks, our HHAs explained they could not be certain whether their reports were successfully sent, since the system did not provide confirmation of receipt. In fact, our HHAs explained they did not trust the punch-code system, because it *"could be bugging"* and fail to deliver their daily reports to their agency.

In these scenarios, protocol dictated that HHAs' agency coordinators were to request that they submit paper timesheets, a process that would require HHAs to travel to agency offices to submit hard copies or fax them within two days of the unreported shift. HHAs who were not able to complete these submissions within the provided timeframe were not paid for their work.

We note the paper submission process is clearly onerous and places the burden of effort solely on the HHA. Nevertheless, participants described the punch-code system to be so unreliable that they carried paper timesheets around with them just in case. One reported submitting timesheets at least once a week:

> *"The system was made so long ago, it doesn't work that good. I know at least once a week I'm going to have to put in one or two [timesheets]. Sometimes I get a good week and I don't have to put in any. Last week I had to put in two."* (P34, female, HHA)

In addition to describing the punch-code system as unreliable and barely usable, HHAs also reported it was not flexible enough to capture vital information, including details relevant to heart failure care. For example, a HHA preparing a meal for a heart failure patient might want to indicate whether the meal had adhered to the low-salt diet often prescribed for heart failure recovery — but such a task could only be captured by a general *"meal preparation"* record under the predefined terms of the task recording scheme.

Lack of care team integration

Participants also described ways in which HHAs' information systems encoded a sharp divide between a patient's clinical and home care teams. We bucket these into two themes: 1) divisions due to privacy laws and 2) divisions due to the culture of medicine.

First, our analysis found that the laws and procedures governing patients' privacy played a central role in the asymmetry of HHAs' technology ecosystem. Specifically, we found that agencies' interpretations of the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA) effectively prevented HHAs from knowing vital information about their patients. This manifested in HHAs' technology ecosystem as a systemic lack of clarity around what exactly their patients were recovering from, relegating HHAs to executing tasks without higher-level rationale.

Under HIPAA, a patient's healthcare professionals (HCPs) were permitted to receive their protected health information (PHI) from other HCPs. However, our analysis showed most agencies did not consider HHAs to be HCPs, and therefore did not always provide HHAs with vital information like a patient's diagnosis or list of medications. As one participant shared:

> *"Home health aides are invisible ... the nurse has the care plan, but because of the HIPAA Law, they [HHAs] can't be informed of certain things ... I think the home care aide is not considered as a professional. So is there any way ... can the HIPAA Law include that change right here, that [HHAs] are entitled to have that information?"* (P4, Female, HHA)

Our multi-stakeholder analysis showed the problem began at the discharge level, when physicians and social workers handed patients off to agencies. One participant shared,

> *"The information we get on the aide side is very skimpy. The initial discharge from the hospital is like 'so and so is going home tomorrow and the pick-up time is at such and such a time.' There's very little information as to what diagnosis the patient has or what medications they are on. Really then the aides are stuck with what to do."* (P31, Female, Nurse)

The effective ban on HHAs receiving PHI had clear ramifications for the quality of care HHAs were able to provide: For instance, a HHA who knew she was caring for a patient with heart failure might emphasize hydration and low-sodium meals more than a HHA who did not. Indeed, participants agreed that the effective ban presented significant obstacles to positive patient outcomes. Social worker participants corroborated HHAs' impressions that initial discharge instructions were left intentionally vague due to HIPAA, and added that this had consequences for patient care:

> *"We work to make the safest plan possible... but because of HIPAA, it's unlikely the person who is with the patient all day ever sees this information. And then the patient comes back, gets readmitted, and the whole process starts all over again... A lot of the time they are in the ER when they don't need to be. I think if we were communicating with the person in the home, providing the care, we would be in a lot better shape in terms of improving outcomes."* (P26, Female, Social Worker)

Working with their patients in their homes day in and day out, HHAs who wanted to know what to expect from their patients' health were relegated to two workarounds: 1) hoping their patients and their families would volunteer the information or 2) inferring disease status from the care plan. Both avenues are far from foolproof, offering ample opportunity for patients and families to misinform HHAs, or for HHAs to themselves erroneously assume what patients are suffering from. We note that several of the more experienced HHAs in our study described they had over time picked up homegrown knowledge on the signs and symptoms of heart failure, and could infer with reasonable certainty when a patient was suffering from it.

We note the centrality of the paper-based care plan was particularly problematic. Participants made reference to state and local health board laws that held patients must be provided with a hard copy of their care plans. While agencies may have intended for these paper-only documents to serve as the definitive record of the tasks HHAs were to perform, our HHAs agreed that many patients hid, obscured, or altered their care plans instead, to the detriment of HHAs' ability to give proper care.

Indeed, stakeholders in our data agreed that the problem was *"deep-rooted"*, affecting agency nurses' ability to provide meaningful instructions on their care plans as well. Nurse participants shared,

> *"We walk in blind. There's a primary diagnosis that is the most recent diagnosis that the plan or the peer source is sending us. There's one ICD-10 code that's written on that authorization. And so you walk in, and maybe you thought it was all arthritis, but [you're actually] ten diagnoses later – which oftentimes you are figuring out from the medications, because the patient may not be an accurate historian. And then you have to validate that with the physician. And that doesn't happen necessarily at the point [where] you've developed the first plan of care with that first aide."* (P3, Female,

> Nurse)

HIPAA also impacted HHAs' technology ecosystem as an effective ban on the use of personal devices in their work for messaging, taking pictures, or otherwise handling patients' information on unsecured machines. When asked whether HHAs engaged in informal communication via consumer chat tools like WhatsApp, HHAs agreed this did not happen *"because of HIPAA"*; agency leaders also agreed blanket bans on the use of personal devices in patient care were the industry norm. As previously noted, however, HHAs in our data admitted the use of personal phones was commonplace in practice, to store patient medication lists or compile photographic evidence of patients' conditions.

Interestingly, our analysis showed that while HIPAA stipulations gave explicit legal definition to some parts of the misalignment, for example by stipulating who could have access to a patient's diagnosis, much of the entrenched asymmetry manifested in the information system was simply cultural. A coordinator said:

> *"It's not necessarily HIPAA. I think most people just feel like a lot of the medical stuff is above the home health aide, so the perception affects what they are then told."* (P29, Female, Coordinator)

Across stakeholder groups, our participants agreed that home care was undervalued by physicians, but vital to patients' compliance and eventual recovery. Indeed, our HHAs felt that even though they served as the *"eyes and ears"* of the clinical team because they spent the most time with patients, their voices were not taken into account when making clinical decisions. We heard of several cases in which doctors and nurses were unaware outright of the activities that HHAs performed during home visits. One participant shared,

> *"We've been doing this work for a long, long time and I've always believed that 'Without a good aide, you don't have a good chance'. You can quote me on that! Unfortunately, because of the way the system is set up – different nurses, seeing different doctors, aides sit outside in the doctors' office – patients are non-compliant. Physicians are just not very involved in home care at all. They don't want to be bothered. The system is fragmented and broken. It's a mess."* (P31, Female, HHA and Coordinator)

The lack of integration manifested in HHAs' technology ecosystem most clearly through the complicated call chain required for HHAs to report incidents to clinicians. Our analysis also showed it led to an overall dissatisfaction among HHAs with the demands of their data collection and reporting tasks. Specifically, participants felt a sense of disconnection in this work, as it was unclear who saw the collected data, and how or in what part the collected data was used by other stakeholders. One participant described,

> *"I think that one thing that was not clear was how [the] data was being managed. There was still a barrier with, is it going directly to the nurse, or what was ... How is it being reviewed? It's all on how the home care aide is instructed on collecting the data... Those kinds of things should be clear on how the aide is documented and the frequency of the documentation, because those things can leave room for what's not important and what is important."* (P2, Female, Nurse)

Sustaining Technology in the Home Care Environment

Our analysis also showed how many of the issues faced by HHAs in their day-to-day interactions with technology could be explained by a lagging level of investment in creating technological solutions that would be sustainable over the long term.

First, our participants were quick to point out that the telephonic punch-code system had been developed and initially deployed decades ago, in an analog technology environment. This legacy system remains in place despite advances in digital technologies in large part because the system relies on HHAs' use of patients' in-home technology infrastructure that, especially in light of the advanced age of many patients, is not guaranteed to be any more updated than analog (i.e., landline) phones.

In addition, although Internet and cellular networking has improved since, many home care contexts lack guaranteed persistent connectivity, even in New York City. One way to resolve this problem would be to set up reliable connectivity that works even in poor cellphone reception. However, agencies had attempted this approach and found the process challenging to implement. One participant explained,

> *"We're exploring external modems with data, without data. It's really cumbersome. It's an interesting thing, because if you would have gone back 10 years ago and said, oh, everybody's gonna get a tablet and be able to connect, it would be a miracle. But it's very cumbersome, and very glitchy."*
> (P17, Female, Agency Leader)

In addition to the technical challenge of low connectivity, our analysis found agencies also struggled with the general challenges of sustaining long-term interventions in community health. Our participants explained that the integration of new tools into existing care workflows often incurred compounding costs. For instance, uptake of a new tool in an agency required not just investing in the tools themselves, but also retraining existing HHAs to use these tools, revamping onboarding processes for new HHAs to include the tool, and establishing new protocols for continuing certification.

Participants explained the compounding costs problem was exacerbated by the transient nature of many HHAs' roles. *"Aide turnovers"* disincentivized agencies from supporting expensive workforce retraining programs, but also made HHAs themselves less likely to participate. Even those HHAs who were open to technological innovation would be reluctant, participants described, to train on a new tool for a job they planned to soon leave.

Finally, one agency leader explained that agencies struggled to deploy new technologies because they required strong partnerships with not just clinical groups but also software companies. Some technological pilots had failed in the past, agency leadership participants explained, due in part to poor collaboration between home care agencies and the software organizations involved. Participants described these partnerships as tough to build because they *"take time and trust"*.

4.4.3 Stakeholders' Suggestions for Technologies

Our study participants were eager to share ideas on how technology could be used to better serve HHAs' needs. From their suggestions our analysis derived three key themes: 1) revising the task recording system to become more flexible and dynamic; 2) enabling communication with clinical teams; and 3) enhancing agency management systems. We detail these in turn.

Revising the task recording system

Across roles, stakeholders in our dataset had numerous suggestions for improving the current task recording paradigm through technology. Acknowledging that a system centered on a static paper-based care plan was ill-suited for home care of heart failure in particular, participants felt the system could be revised to emphasize *flexibility* beyond the care plan schema. HHAs voiced that they often wanted to record more finely grained data on a patient's mood throughout a shift, whether there had been any changes in appetite, and other pertinent information that existing care plan-to-punch code mappings did not account for. Participants also agreed that adding richer media to task records would improve their experiences, in particular photos and videos taken with a mobile device.

In addition to expanding the schema and data types used in task records, participants saw a need to revise the entire task recording framework from a *post facto* "clock-out" procedure built for billing purposes to a dynamic "just-in-time" system built to supply HHAs with decision support at the point of need. Such a system could, for example, send a HHA tips for low-sodium cooking as he or she was preparing a meal for a patient with heart failure. This was viewed as one way to honor the clinical significance of the care plan while enabling dynamism in HHAs' experiences through technology.

Enabling communication with clinical teams

Our analysis also found that improving communication along both the HHA-agency nurse and HHA-doctor axes would be a way for technology to improve HHAs' experiences. As described previously, stakeholders agreed that HHAs needed on-call access to a verifiable clinical opinion. One participant said:

> *" 'Here's these three things, and there's gonna be a nurse on call who can answer your questions if any of these three things happen today.' The experience for the aide, and the outcome for the patient, would be changed dramatically if that could be universal."* (P5, Female, Agency leadership)

Despite widespread agreement that providing HHAs with access to clinical opinions would improve both system inefficiencies and patient care, participants across roles also agreed that a direct-access system could result in *"a fire hose of random messages"* that might overwhelm clinicians' already-stretched time. For perspective, one nurse in our research was in charge of 66 HHAs. To mitigate potential communication overload, our participants recommended a level of filtering for "urgent vs. non-urgent" issues. Clinicians also suggested a role for technology in ensuring they were only contacted if patient metrics exceeded set thresholds. One shared:

> *"You know, if I was getting loads of info from home health aides, regularly, that would be annoying. But if it was a once in a while alert of a change, now that would be potentially quite useful. It would depend on how it was done."* (P28, Female, Physician)

Participants also agreed that direct communication tools could be difficult to implement given the previously discussed restrictions on HHA access to PHI under HIPAA. However, our analysis found stakeholders across roles appeared flexible with respect to allowing HHAs more access to patient information that would help them provide high-quality care.

In addition, HHAs had developed their own mental models for which care tasks might correspond to which conditions—for instance, mapping low-sodium diets to heart failure—and desired clear clinical education to supplement what they had learned through experience. Most tellingly, patient participants in our study described wanting HHAs to have access to their PHI in electronic form, in order to improve their quality of care. One patient participant shared:

> *"Well, give [my aide] access to my chart, and let them know what maybe thoughts are with blood work and stuff, so they have an idea what's going*

> *on with me. I want them to have access to all of that. I wouldn't want to keep it private because they're helping me to get as better as I'm going to get."* (P21, Female, Patient)

Enhancing agency management systems

Finally, our analysis showed key opportunities for technological innovation in the data management systems used at the agency level. Our agency leader participants were frustrated by the user experience of existing agency management software systems because these systems did not provide the necessary affordances for robust interaction with reported data. For example, agencies relied on a software tool that involved multiple labor-intensive steps that could have been automated. One participant explained,

> *"We're all sort of brought onto a software package, which is a piece of dreck. You know what dreck is? Dreck is garbage, that's all it is. Nothing worse than that... To give you an idea, I'm currently compiling and managing a list of all of our patients who receive Medicaid home care services. I have to update that every week by asking. There's no place for me to look."* (P17, Female, Agency Leader)

Although one could manage by hand a data system for a small set of users, this approach would not be feasible for agencies in our research context that consisted of nurses, coordinators, and hundreds of HHAs. As a result, participants desired a system that went beyond *"basic"* functionalities to provide more relevant features, for example the ability to re-organize documents for easy retrieval instead of simply uploading them. More importantly, they wanted these features available without becoming *"very glitchy, very labor-intensive."*

In addition, our participants yearned for a system that simplified the experience of assigning HHAs to patients. Recall we previously described that HHAs could work for more than one agency and that multiple HHAs could be assigned to care for a single patient. As such, it is paramount for agencies to understand the schedules and availability of patients and HHAs to successfully pair them. However, existing technological systems were lacking in this regard as they only provided a view of patients' schedules but not that of HHAs. One participant described,

> *"It's fundamentally flawed. A HHA is supposed to come in for in-service, supposed to go to get their physicals done. You can't schedule the aide when they've got that work being done, but you will never see it. It's a very clunky system, and then the work in the field is just awful... If I'm a HHA who works for two agencies . . . the software package doesn't differentiate between one and the other... They haven't figured it out yet."* (P16, Male, Agency Leader)

This lack of cross-agency scheduling integration for both HHAs and patients is problematic because it leads to schedule conflicts: HHAs who appear available to one agency may not be truly available because they have been assigned by other agencies to care for other patients. Unless agencies have the opportunity to *"to look at schedules in two ways"* i.e. for both HHAs and patients, the experience of assigning HHAs to patients may become increasingly unproductive with time spent rescheduling conflicts.

Agency leadership participants also voiced a need for better integration of their backend data. It was challenging, agency leadership reported, to retrieve and combine patient data recorded during home visits and hospital discharge because these data were stored in siloed storage systems. One agency found this problem so challenging that it felt data integration could not be achieved; the agency consulted a third-party company to assist but the company also struggled to accomplish this task:

> *"We have five software packages that lead to 10 disparate databases, and those databases are clunky... We actually engaged a third-party software company to come in and pull all the data out of those systems, normalize it, pull it into a data warehouse... And we had a company, a national company that said they wanted to do that and work with us and partner with us, and at the end of the day, we just heard, this past week, they just can't do it. The work effort is just too heavy a lift, they can't do it."* (P16, Male, Agency Leader)

However, data integration is complicated because it involves several non-trivial steps that individually require a lot of time and effort to accomplish. These steps include but are not limited to: tracking down all the data reporting systems used by different health institutions, gaining authorizations to access the data in these systems, understanding the format of data stored, normalizing the data to a uniform format that allows for integration, and eventually integrating the data in a digestible format for agencies to access.

In addition, participants believed that resolving the challenge of data integration could provide rich opportunities to combine data from multiple sources and gain deep insights on how to improve the quality of care provided to patients. One participant shared,

> *"The more information we gather through assessment and clinical and non-clinical documentation, we can pull into this system and then identify people that are truly at risk . . . Like if a social worker does an environmental scan on the home, and realizes something's not right, and is documenting it*

> *on a record somewhere, it would send an alert to our population health unit and say there's a potential risk of fall. And then they could be in communication and do coordination internally on behalf of the patient."* (P16, Male, Agency Leader)

As envisioned in this quote, fusing diverse types of data from multiple sources on the backend could work in tandem with a *"smart"* notification system on the frontend that informs the actions of all providers caring for patients.

Finally, agency leadership participants noted that as of January 2019, U.S. federal law required home care visits to be electronically verified per the 21st Century Cures Act. This legal mandate, participants noted, laid the groundwork for successful partnerships between software companies and home care agencies that might make sweeping changes possible, and even essential.

4.5 Discussion of Qualitative Fieldwork

Having developed a nuanced understanding of how technology influences the way HHAs care for adults with heart failure, we now synthesize our findings into design opportunities for HCI researchers, system designers, and medical researchers interested in developing technologies that support formal caregiving for heart failure patients. In addition, our paper answers an important call for additional research on the role of HHAs in heart failure self-care [229] and we provide concrete suggestions for future directions.

Knowledge Gaps in Decision Making

Our analysis highlighted that many HHAs do not always have a clear understanding of what to do when patients' conditions change, resulting in unwarranted emergency calls to 911 as knee-jerk reflexes. Moreover, HHAs are at-tempting to fill their knowledge gaps by Googling for relevant information. Prior research revealed that HHAs find heart failure highly unpredictable and frightening [231, 230], which could in part explain the high rate of 911 calls from HHAs covering their bases. Our work builds on these previous findings and suggests the need for an improved system that supports HHAs in making informed decisions.

A key design opportunity that could reduce knee-jerk reactions and bridge HHAs' knowledge gaps might be an effort to educate HHAs on heart failure symptoms and specific scenarios that require emergency calls. This educational content could be provided to HHAs as static digital information that is accessible in offline mode on a digital device (e.g. a tablet). Our finding that stakeholders saw a need for *"just-in-time"* supplementary information on heart failure also suggests a role for real-time instruction from a technology akin to a personal assistant. For example, if a HHA indicates on the tool that a patient has fluid retained in their body, then the tool could recommend *"lift the patient's legs for 30 minutes"*; if it is reported that the patient's weight suddenly increased by two pounds in the last 24 hours and the patient has short breaths, then the tool could show a red alert recommending *"call the nurse now"*, and provide the functionality to do so.

Similar to prior work in the infographic and design communities that adopt pictographic representations to communicate complex information [95, 94], these educational and decision-support tools could break down complex heart failure symptoms and next-step actions into digestible formats.

Adopting these techniques could reduce the level of uncertainty that HHAs face, and potentially decrease the amount of unwarranted emergency calls that occur. Without addressing this challenge, more financial and medical resources could be wasted on false emergency calls. We note that in the U.S., patient readmission within 30 days incurs financial penalties for hospitals due to Medicare programs around value-based purchasing and hospital readmission reduction [83]. On a global scale, reducing unnecessary emergency calls could also contribute to a more sustainable healthcare approach.

Privacy Concerns in Health Technologies

The extent to which U.S. medical information privacy laws were enforced in home care played a big role in our research context. Concerns around privacy surfaced on two fronts: first, in communications between members of a patient's clinical team and the HHA; and second, in the use of HHAs' personal devices to record and transmit information about their patients. Stakeholders across roles voiced that patient privacy was important, and should be respected and encoded throughout the information systems enabling home care work. However, participants also agreed HIPAA's effective exclusion of HHAs from receiving vital patient information was significantly limiting the quality of home care. This finding adds to the growing amount of research on the tension between government regulations and the implementation of health care systems [132, 5]. We note this is especially pressing in our research context because it involves heart failure, a life-and-death disease.

One design opportunity to address this tension could be to *"build a long lasting relationship based on mutual trust and respect"* by adopting the principles of privacy by design [144] in the implementation of technological systems for heart failure management. An agency could, for example, provide tablets with task recording tools and calling features directly to patients as part of their services. These tablets would be secured at patients' homes for rotating HHAs to use for 'clock in and out'. The use of dedicated devices that remain at patient sites would enable these systems to comply with the requirements of HIPAA, including password protection, data encryption, access control and logging [120]. Becoming HIPAA compliant would also enable agencies to provide HHAs with patient information vital to their work, for example patient diagnoses. Adopting these processes could assist stakeholders in reducing ongoing tension between agencies, HHAs and heart failure patients.

Partial Reporting of Patient Conditions

A key finding in our research is that restrictions in reporting tools may prevent HHAs from communicating when patients are in life-threatening conditions. After an episode of home care, HHAs are required to use their patients' phones to call an automated punch-code system to report all the activities they performed. During the call, the HHA has to scan through a long list of tasks on a coded sheet to find all the activities they performed, and for each activity punch the corresponding numerical value using the phone keypad. For example, *cooking a meal* could have *99* as its assigned value, so the HHA types *99* during the call. If a phone call cannot be placed due to poor cellular network in the patient's home, the HHA records all activities performed on a paper task sheet. HHAs who do not follow these procedures risk losing their jobs.

However, our work highlights several challenges with this reporting process, especially in heart failure. As a disease characterized by an unpredictable and uncertain symptom course, heart failure often requires that all important activities performed for and with patients be captured. For example, HHAs need to monitor patients' salt intake as this could lead to a serious condition of fluid retention and swelling of the body; but list of codes baked into the current industry-standard system does not allow the reporting of *"salt intake"*, and instead broadly captures *"meal preparation"*.

In addition, if HHAs erroneously *"punch"* in the wrong code during reporting, they cannot undo their actions. HHAs do not have the opportunity to review their entries or receive confirmation after submissions have been made. In cases where paper reports are used, HHAs need to fax or submit reports in-person on time or they will not be paid. These drawbacks suggest that HHAs need a more effective process that accommodates the reporting of heart failure activities, allows updates of incorrect entries, and remains fully functional during poor network connectivity.

One design avenue for an effective reporting process could be to create an offline-first tool for HHAs to collect and report data in a modular and granular fashion. The tool could remain fully functional when there is poor network connectivity and then sync back to the cloud when a reliable internet connection is established. This offline-based approach has been well studied and successfully implemented in tools (e.g., Medic Mobile [169], ODK [32], and CommCare [65]) for frontline health workers across the globe.

In addition, the reporting tool could be designed in form of a mobile or web application where HHAs can indicate the specific care activity they performed by quickly clicking through a form that contains both general activities (e.g. laundry) and heart-failure specific tasks (e.g. prepared a low-salt diet). To reduce errors in reporting, the tool could provide an opportunity to review all entries before submission and update activities erroneously logged. Adopting a more streamlined approach of seamlessly capturing care activities during home visits could improve the quality of care provided to heart failure patients.

Promoting HHAs' Value and Integration

Stakeholders across roles expressed a need for greater integration of HHAs into patients' broader clinical teams. Downstream from the finding that stakeholders would like for HHAs to have access to patient diagnoses, we see opportunities for systems to better integrate HHAs' perspectives and the data they collect. This was critical for HHAs in our study, who felt marginalized — they labored to collect data but lacked visibility into its context and use, and were not respected by the system at large. Agencies similarly felt that the data collected by HHAs were spread across several isolated systems in ways that made it complicated to retrieve and perform integrated analyses that might improve patient care. Health practitioners worried that they were missing vital information into patients' everyday conditions – information to which HHAs had access – and simultaneously worried that tools that kept them updated in real time may become overwhelming.

The need for thoughtful integration of HHA perspectives into the broader team presents an opportunity for the design of an end-to-end information system supplying stakeholders with detailed information on configurable notification cycles. On a data level, the ability to see trends in the data they had collected over time might alleviate HHAs' yearning for context. Doctors and nurses might also have their needs addressed through a system allowing them to set notification thresholds and cycles for key patient metrics, for example enabling them to stipulate they only wanted alerts-of-change emailed on a weekly basis. A robust system centered on the collection of patient datapoints relevant to home care could provide agency leadership with the insights they need to optimize outcomes, and scheduling functionalities, strictly enforced, could help agencies achieve operational efficiencies.

On a communication level, stakeholders' desires for better communication might be addressed through a system enabling HHAs to triage whether a patient's condition merited an emergency services call, a call to an on-call nurse, or simply an asynchronous message notifying a patient's doctor of a change. Blending ideas from clinical decision support and call center escalation protocols, such a system might also provide HHAs, coordinators, nurses, and other members of the call chain with *visibility* into the the status of their inquiries (e.g. "Forwarded to primary medical doctor" vs. "Forwarded to agency coordinator").

These improvements would directly address the immediate needs of stakeholders laboring in this asymmetric information environment. While achieving parity for HHAs would require systemic cultural shifts, for which the structure of the underlying technologies can only do so much, we see design opportunities for information systems to encode new social paradigms. For example, we envision compelling future work exploring whether and how algorithmic decision support and triage might change perceptions of what patient states require emergency or medical attention — among not just HHAs, but also doctors, nurses, and patients.

4.6 Part II: Creating a Digital Provocation

Building on the insights from our qualitative research, we created a digital provocation that engages with the challenges and tensions identified. Then we tested it with stakeholders working within the ecosystem to understand their reactions. In this section, we begin by discussing 1) how we arrived at our design provocation methodology. We then 2) detail specific design choices we made in the course of designing the provocation, and 3) describe our field study with 16 stakeholders across aide, nurse, and coordinator groups.

4.6.1 Design Method

Methodologically, we were inspired by ongoing work within the interaction design and HCI communities advancing discursive design methods such as speculative design [4] and design fiction [25, 151, 152, 228]. Our goal was not to "solve problems" for our users (an approach Vallgrda et al. [248] characterize as movement towards a defined solution within a known context) but rather to use design to explore the space of potential contexts. We contrast this approach with traditional methods in user research, i.e. interviewing or contextual enquiry, which are more focused on arriving at a nuanced understanding of an existing design space. Thus we consider the technology design in our study not as a *prototype of a tool* but rather a *provocation*, and focus our analysis on participants' *interpretations* of the design.

Specifically, we sought to present stakeholders with a deliberately incomplete vision of a future in which aides are more integral to the healthcare team and have more "clinical" responsibilities—an outcome increasingly possible under the tectonic shifts in healthcare payment models in the U.S. towards value-based payment schemes. Our provocation was created to seed this broad vision with specific cues, while refraining from prescriptively stifling participants' imaginations: a *medium*-fidelity prototype of a digital application with some level of interaction, but without actual function. By centering this digital technology on the aide as its sole user, we aimed to create a material starting point from which our participants could, through reactions to the provocation, begin to articulate aides' experiences as well as their visions of the future.

4.6.2 Designing the Provocation

Recall that in our prior qualitative research, synthesized **three key needs** for aides: 1) the ability to record their daily tasks seamlessly and reliably; 2) the ability to communicate with agency nurses and supervisors in real-time, including flagging emergencies; and 3) the ability to access on-device educational content on medical information, like normal blood pressure ranges, and health information, such as the components of a low-salt diet.

We used these needs as a starting point for our provocation and constructed a medium-fidelity prototype of a tablet application that appeared to fulfill them. We chose the tablet form factor because it offered a mobile device modality familiar to participants, with enough flexibility to plausibly encapsulate data entry, communication, and multimedia educational content. We created a series of screens that mimicked the design motifs expected in a modern mobile application and that enabled users to tap on select buttons to move from screen to screen. However, we stopped short of actual functionality in order to leave space for future imaginings. For instance, the provocation does not store information or realize any actual real-time messaging. We now discuss how our provocation engaged with aides' three key need areas.

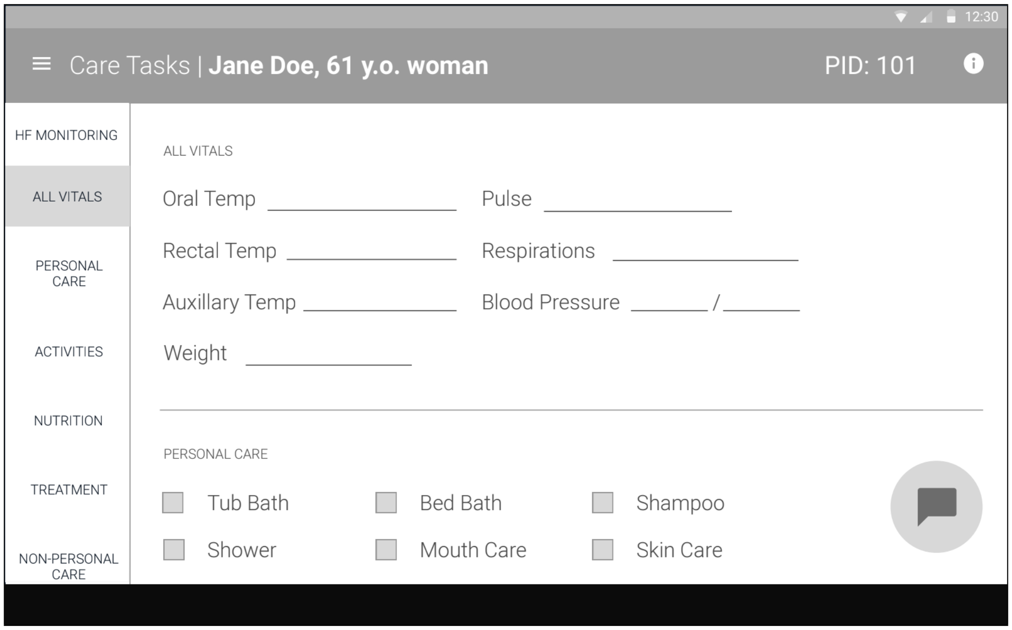
The ability to record daily tasks seamlessly and reliably. As previously described, a home health aide's work revolves around recording the tasks they have completed for their clients via a telephonic punch-code system. Each day, when an aide arrives at her client's home, she uses the client's home telephone to call into an automated phone line. This is how aides ”clock in”, or register with their employers that they have arrived at the client site.

The aide then consults the client's *care plan*, a document outlining the agreed-upon tasks the aide is to perform for the client that day. Care plans are often left in paper format for the client to display prominently in the home, for example on the refrigerator. The aide goes about executing against the care plan throughout her shift.

At the end of her shift, the aide must dial again into the automated telephone system using her client's phone and follow a series of prompts to ”clock out”. As the system records her time of clock-out, she also completes an important step: task recording via telephonic punch-code. Aides enter numbers into the system that map to the activities they performed for the client that day. This registry of aides' activities is critical to insurance companies' reimbursement models for home care, and forms the basis of agencies' electronic visit verification (EVV) schemes.

Despite occupying a central space in how aides interacted with their agencies, these systems are often seen as unreliable and cumbersome. In fact, they are often the source of tension between aides and their coordinators. In the event that the telephonic punch-code system fails to record an arrival, the latter has to call the patient to verify that the aide had indeed made a visit. As previously described, this has shown that aides have so little trust in the existing technology that they carry paper timesheets just to be safe. The telephonic punch-code system also offers users no visibility into the numbers they have already entered, and no way to correct an entry in the event that a user accidentally enters the code for a task they did not perform.

Our provocation engages with the challenges described above by appearing to provide aides with digital tools for the data collection they already conduct. Specifically, we constructed a way for aides to 1) reference clients' paper care plans in digital format by viewing a scanned image of the paper care plan, and 2)



ALL VITALS

Oral Temp ____________ Pulse ____________

Rectal Temp ____________ Respirations ____________

Auxillary Temp ____________ Blood Pressure ______ / ______

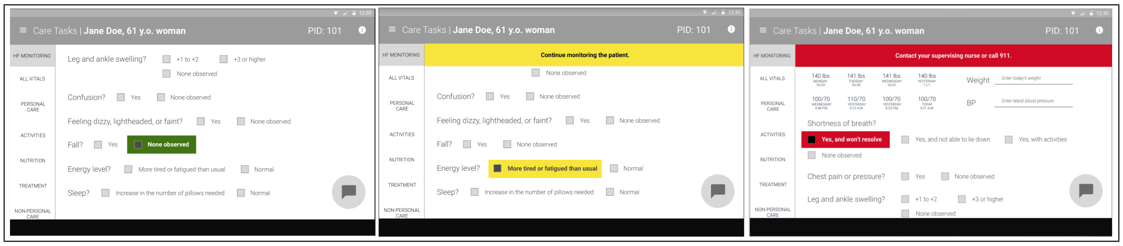
Weight ____________

PERSONAL CARE

☐ Tub Bath ☐ Bed Bath ☐ Shampoo

☐ Shower ☐ Mouth Care ☐ Skin Care

Figure 4.1: Design Provocation: Interface for recording tasks.



Leg and ankle swelling? ☐ +1 to +2 ☐ +3 or higher ☐ None observed

Confusion? ☐ Yes ☐ None observed

Feeling dizzy, lightheaded, or faint? ☐ Yes ☐ None observed

Fall? ☐ Yes ☒ **None observed**

Energy level? ☐ More tired or fatigued than usual ☐ Normal

Sleep? ☐ Increase in the number of pillows needed ☐ Normal

Continue monitoring the patient.

☐ None observed

Confusion? ☐ Yes ☐ None observed

Feeling dizzy, lightheaded, or faint? ☐ Yes ☐ None observed

Fall? ☐ Yes ☐ None observed

Energy level? ☒ **More tired or fatigued than usual** ☐ Normal

Sleep? ☐ Increase in the number of pillows needed ☐ Normal

Contact your supervising nurse or call 911.

140 lbs MONDAY 10/29	141 lbs TUESDAY 10/30	141 lbs WEDNESDAY 10/31	140 lbs YESTERDAY 11/1	Weight *Enter today's weight*
100/70 WEDNESDAY 4:58 P.M.	110/70 YESTERDAY 9:12 A.M.	100/70 YESTERDAY 5:23 P.M.	100/70 TODAY 9:21 A.M.	BP *Enter latest blood pressure*

Shortness of breath?

☒ **Yes, and won't resolve** ☐ Yes, and not able to lie down ☐ Yes, with activities ☐ None observed

Chest pain or pressure? ☐ Yes ☐ None observed

Leg and ankle swelling? ☐ +1 to +2 ☐ +3 or higher ☐ None observed

Figure 4.2: Design Provocation: heart failure-related decision support: Green: no cause for concern at this time; Yellow: cause for concern, keep monitoring the client; Red: possible emergency, seek help.

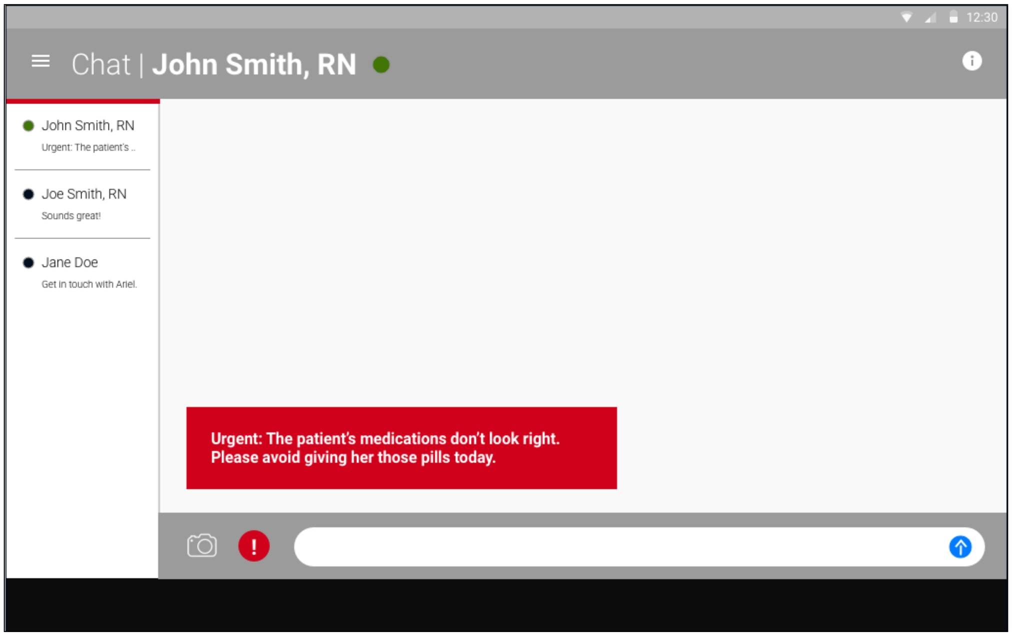
record their daily tasks in an interface that appeared to provide reliability and convenience (Figure 4.1), all on an application on a tablet provided for aides' use.

Our provocation also extends the themes of this need into newer territory: What if at the times of clock-in and clock-out aides also collected and uploaded data on their client's health status? For example, an aide who helped her client keep track of her weight over time could record that she had completed this task, and she could also record those values and send them directly to the client's clinical team. We saw this as an avenue through which aides might achieve elevated importance within home care, since aides are the natural observers of data to which clients' doctors and nurses currently do not have access.

We also saw the recording of patient medical indicators as an opportunity to provide aides with "just-in-time" decision support as aides often struggle with whether and when to call for help from emergency services or from a client's doctor. To extend the possibility of aides making more informed decisions in these moments of crisis, we considered the question: What if aides received immediate feedback from a technology tool itself on whether clients' levels of leg swelling were cause for alarm, and when to call a doctor? We saw such decision support as an avenue for aides to be equipped to provide better and more efficient care to their clients.

Figure 4.2 shows the screen of our provocation that realizes these ideas. Care tasks like personal care and nutrition are tabbed on the left-hand side of the interface. A tab called "HF Monitoring" shows inputs where aides might enter medical indicators with clear relevance to heart failure, such as weight and blood pressure, as well as checkboxes for additional relevant indicators like shortness of breath, chest pain, and leg or ankle swelling. As shown, an aide who selects an indicator that gives reason for alarm is directed to contact a supervising nurse or 911. Of note, in developing the decision support feature of our provocation, we consulted with nurses and medical doctors with experience in home care of heart failure patients for a clear decision-making algorithm aides might use.

The ability to communicate with nurses and coordinators. Much of aides' current daily work also revolves around communication with clients' clinical teams and with agency coordinators. Coordinators are full-time office-based employees who work on rotating shifts, are aides' immediate supervisors at home care agencies. They are responsible for assigning aides to clients, directing



John Smith, RN
Urgent: The patient's ..

Joe Smith, RN
Sounds great!

Jane Doe
Get in touch with Ariel.

Urgent: The patient's medications don't look right.
Please avoid giving her those pills today.

Figure 4.3: Design Provocation: the instant message feature.

them to client sites and monitoring aides' clock-in, clock-out, and task recording data for inconsistencies, such as a failure of the punch-code system to record a visit, which requires them to call the client. Clients and their family members also often correspond with coordinators to manage their arrangements with the agency.

Nurses also work for an agency on rotating shifts. When a client is deemed to be home-bound and have skilled need, a nurse is assigned to their care. The nurse then provides an initial assessment, drafts the care plan, and subsequently supervises the aide (who provides the day-to-day care). Nurses then visit every few weeks to check on the client and adjust the care plan as needed.

While coordinators are meant to be aides' first phone calls for all issues that arise while providing care, aides have voiced it is often difficult to contact coordinators in a timely manner and to contact nurses directly when they need clinical guidance and when coordinators, who do not have medical expertise, are not equipped to help. Currently, communication largely relies on phone calls, and coordinators and nurses alike may manage dozens of aides at once. As a result, aides can often spend hours waiting for nurses and coordinators to return their call.

We engage with these challenges in our provocation through a chat feature that appears to provide aides with an interface for instant-messaging nurses and/or other agency staff (Figure 4.3). This alone suggests disruptions to the current workflow, moving communication from phone calls to asynchronous messaging and providing aides with a direct line to clinical staff. This chat interface also appears to be accessible from all other screens in our provocation as a persistent chat icon. On several screens we also created the appearance of new message notifications with a red flag on the chat icon.

To extend the disruptive potential of our provocation, we suggest additional possibilities for this computer-mediated communication channel, through design cues. A green circle next to the nurse's name hints at whether the nurse is online or offline. A red icon next to the message input box suggests the ability to flag certain messages as urgent. Finally, a camera icon in that same dialog suggests the ability to send photos or videos. In order to maintain focus on future communication possibilities, we did not incorporate explicit cues to signal whether the messages exchanged were compliant with Health Insurance Portability and Accountability Act (HIPAA), a U.S. statute that sets privacy guidelines for patient information.

The ability to review on-device educational content. Aides working with heart failure patients have expressed that they often feel the need for more health and medical education to properly care for their clients [229, 230]. Agencies currently provide required training programs (known as ”in-services”) that might cover relevant baseline information, like the elements of a low-salt diet

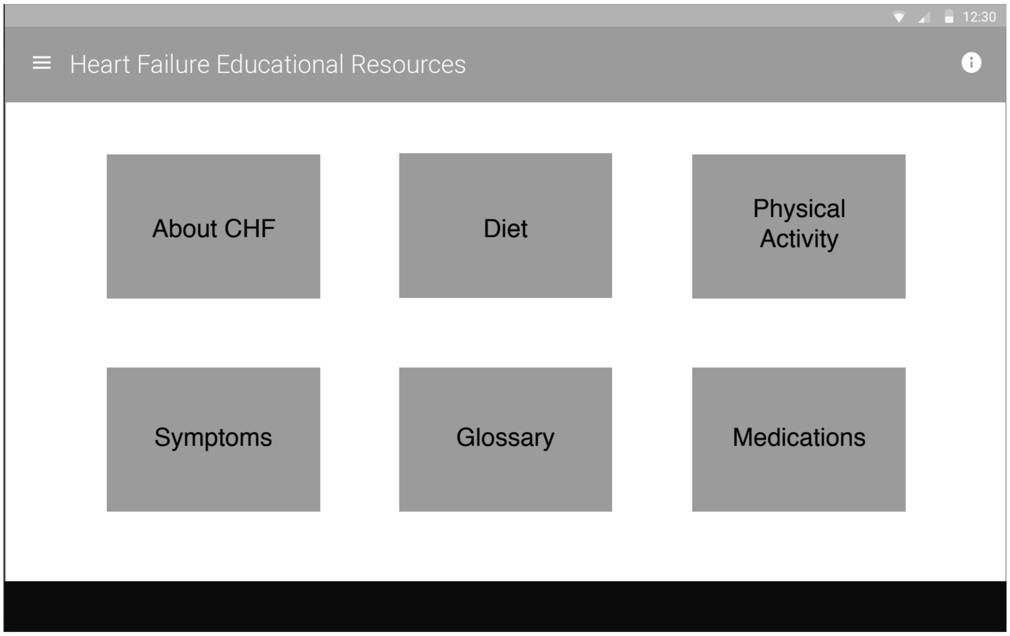


Figure 4.4: Design Provocation: heart failure education screen.

or tips for keeping a patient's legs elevated, but as described earlier, aides supplement these trainings with information they source on their own, either from other aides or by searching the Internet.

In our provocation, we engaged with the need for educational resources through a button on the "home screen" labeled "*Learn about Heart Failure*." As shown in Figure 4.4, the provocation contained high-level information categories, but it did not provide actual resources. We intentionally kept this section of our provocation vague, to leave room for participants to fill in what it might contain. This section was also developed in consultation with nurses and medical doctors familiar with heart failure care.

Incorporating a behavioral design lens Since our intervention intended to change the behaviors of home health aides and their existing workflow, we engaged with Behavioral Science literature for techniques that directly addressed decision making, a key challenge for HHAs. In our review we considered *nudge* *theory* [240], *knowledge learning theory* [86] and *simple heuristics* [46]. Although the last two concepts were relevant, they were more appropriate after a heart failure curriculum had been designed for HHAs. Building on *nudge theory*, we included in our provocation as previously described, the immediate feedback responses delivered in different background colors that were either *green*, *yellow* or *red* alert signals, as a way to guide HHAs in their work. For example, suppose an aide is reporting their daily tasks in the tablet and they indicate that a patient fell but still looked fine, the app would display a message in *red* background prompting: *"contact your supervising nurse immediately"*. By designing the provocation based on behavioral theories, we aspired to make it more a more persuasive tool for aides.

4.6.3 Field Study Testing the Provocation

We conducted a field study in the winter of 2018 and spring of 2019, in which we showed our provocation to 16 participants from three stakeholder groups (Table 4.2), including aides, nurses, and coordinators who worked with multiple agencies. Our research team consisted of three women and one man who all reside in the U.S. All had experience working with underserved populations. Three have extensive research experience designing technologies for marginalized populations in low-resource environments. One researcher, a medical doctor, has clinical expertise caring for heart failure patients at a large academic medical center and established relationships with multiple home care agencies in New York state. We received IRB approval for all study procedures.

Recruitment. We recruited participants through direct outreach to home care agencies and snowball sampling. Since they had been recruited through their employers, participants were assured that participating in our research would not affect their employment status or benefits, and that participation would be strictly anonymous. All participants provided written or verbal consent to participate.

Participants. The majority of our participants were women (see Table 4.2) who worked primarily in English. A notable proportion of our aide participants (4/11) worked at least half the time in Spanish.

In Table 4.2, Agency A is a worker-owned home care agency headquartered in the Bronx, NY. Agency B is a community-based nonprofit in Queens, NY. Agencies represented with ”Other” are all comparable home care service agencies in New York, NY. All agencies in our study follow many of the same industry-standard procedures, including having aides utilize the telephonic punch-code system for clocking in and out, routing aide communications through coordinators, and centering day-to-day operations around the plans of care created by traveling agency nurses and left with the client. All agencies also operate under the jurisdiction of the Health Insurance Portability and Accountability Act (HIPAA).

Design Sessions. We conducted eight focus group sessions, with 2 to 3 participants per session, each lasting 60 to 90 minutes. We chose focus groups rather than one-on-one interviews because we wanted participants to generate ideas as a group and to counter-point each other. Though aides were the target users of the provocation, we also solicited the perspectives of nurses and coordinators, because they work closely with aides. All focus groups were homogeneous with respect to stakeholder category, meaning aides were grouped only with other aides, nurses with nurses, etc. All sessions were conducted in English, except

Ppt ID	Role	Employer	Language	Gender
1	HHA	Other	English	Female
2	HHA	Other	English	Female
3	HHA	Other	English	Female
4	HHA	Other	English	Female
5	Nurse	Other	English	Female
6	HHA	Agency A	English	Female
7	HHA	Agency A	English	Female
8	HHA	Agency A	Spanish	Female
9	HHA	Agency A	Spanish	Female
10	HHA	Agency A	Spanish	Female
11	Nurse	Agency B	English	Female
12	Nurse	Agency B	English	Female
13	HHA	Agency B	English	Female
14	HHA	Agency B	Spanish	Female
15	Coordinator	Agency B	English	Male
16	Coordinator	Agency B	English	Female

Table 4.2: Participant demographic characteristics

one conducted in Spanish via an interpreter familiar with the domain.

Each session began with a discussion about participants' roles in the heart failure ecosystem, including the challenges they experienced in their roles. All participants were then shown the provocation (displayed on a real tablet) and were encouraged to interact with it, to freely share reactions and thoughts about how it might impact their work and possible challenges to adoption, and to imagine it in different settings. Drawing on design recommendations of prior work [184] that suggested aides could carry out their work in a HIPAA compliant way by working on a dedicated work device instead of using their personal devices (as seen in current practices), we asked all participants about their perspectives on keeping the tablets in client's home compared to aides carrying the tablets at all times.

Data Analysis Design sessions were audio-recorded (where participants permitted) and professionally transcribed. We also took detailed notes during all sessions. We analyzed this data thematically [234], beginning with a close reading of the data and allowing codes to emerge. Multiple passes through the data resulted in 58 distinct codes (e.g., *magical technology*, *paper trail*, *aides want to learn*, and *distrust of Internet information*). We clustered related codes into high-level themes (e.g., *record-setting*, *mismatched interests*, *triggers conflict*) and organized them in a codebook. After multiple discussions and iteratively refining the codes and themes, we arrived at a final set of themes that comprehensively represented the data.

4.7 Findings from Testing the Provocation

Our findings cover five major themes that emerged from our design sessions. We begin by 1) describing how aides perceived the provocation as a way to gain more control over the narrative of their work. We then 2) highlight a clash in how aides and other stakeholders envision the utility of the provocation's communication functionality. We show that 3) participants across groups worried that the aides' use of tablets in clients' homes could trigger conflicts with clients or families of clients. Next, we 4) discuss how participants' interest in the provocation's educational resources differed sharply between aides and other stakeholders. Finally, we 5) uncover participants' perceptions of how the application would increase learning requirements and workload for aides.

4.7.1 Taking Control of Setting the Record

We initially hypothesized that, given the precedent established by the telephonic punch-code system, aides might perceive the task recording features of our provocation as yet another means through which their employers could monitor whether aides were performing their work and mete out disciplinary measures if aides were not performing up to par. Instead, we found that aides viewed the provocation as a way for them to take control and set the narrative around their work—to actively engage in record-*setting*, instead of passively having records kept on them. This affordance was immediately relevant for many of our aide participants' ongoing relationships with their employers; specifically, the records would help them negotiate conflicts over whether or not they had performed their work. Several aides shared that they were already keeping personal notes of their tasks in separate paper notebooks or digital calendars on their personal smartphones. A system like the provocation, one participant said, would provide an additional point of leverage:

> *"This would protect me, because once [my task] is in, it's recorded. There's no 'you didn't call'. It would be my backup, my paper trail."* (P6, aide)

These records were seen as especially powerful for cases that could veer into medical emergencies (e.g. a patient's feet suddenly becoming swollen—a common scenario in heart failure patients). Aides described that after these sort of situations, they often felt pressure from both their agencies and their clients' broader clinical teams to have taken some "*creative*" action to help the patient, or at least to have called their coordinator to report the situation.

But to aides, the issue was not so much about placing phone calls to their coordinators or to a nurse to seek recommendations on what to do in dire situations; rather, it was that these phone calls often went unanswered. With a tool like our provocation, aides felt they would have a reliable way to establish and prove that they had done everything they could. One participant shared:

> [*"A bed sore progresses. I've made several calls, it's growing, it gets worse and worse. This would document my persistence."* (P7, aide)

As this quote suggests, beyond the informal personal record-keeping they already do, aides perceived that our artifact would further aid them by serving as a formal record of the actions and steps they had taken, demonstrating that they were indeed performing their work. To help them record information in more detail, aides also immediately wanted the design to include features beyond those depicted in the provocation and inquired, for example, if it could be used to take photos and videos as part of record-setting.

4.7.2 Clash in the Desire for Better Communication

A key finding in our work is that aides, nurses, and coordinators perceived the real-time instant messaging component of the provocation differently. On one hand, aides viewed this as a step in the right direction to improve current communication practices and potentially serve as a vital tool for their day-to-day work. Recall that in the previous subsection, aides complained of reaching out by phone to nurses and coordinators, but these calls frequently went unanswered. Upon learning about the chat feature in our provocation, they perceived it as a way to address their ongoing communication challenges. Unlike unanswered phone calls that did not convey when situations required urgent responses, aides felt that urgency flags in our provocation could alert nurses and coordinators whenever aides found themselves in dire situations.

However, nurses, on the other hand, felt the communication functionality would give aides an opportunity to abuse their time and reach out to nurses unnecessarily. Furthermore, while *nurses* saw the urgency flags in our provocation as a potentially useful way to triage within what they anticipated as a *"fire hose"* of messages from aides, *coordinators* felt differently. One coordinator (P15) pointed to the red button on the chat interface that provided a way to add an *urgent* flag to a message, and said, *"If I know these aides, everything would be red."* Indeed, for coordinators the chat features of the provocation sparked many negative reactions. Aides already sought to monopolize their time, coordinators explained, and the addition of another way to reach them would only create more work for them.

Another issue was aides' tendencies to want to reach out to *specific* coordinators, regardless of whether or not that coordinator was currently assigned to their current case. While an agency might have staffed up or down to handle an aide's incessant calls, aides often wanted to speak to a coordinator with whom they had already established trust. This created difficult situations for coordinators in which they were often implicitly expected to be available for an aide long after they had left their case. One coordinator participant, P16, explained that after accidentally giving out her personal cellphone number, she was currently experiencing an influx of communications from aides who wanted to speak to her and her only:

> *"On the weekend, I say I'm not working. They'll call me on my cellphone, because when you use your phone to call them, they get your number. And I say, 'I'm not talking to you, I'm not working. Call the emergency number, because I'm not working today'... If you forget to block your number, it's free-for-all. They WILL text you back on it."* (P16, coordinator)

Our provocation raised these questions of nurse and coordinator availability for participants across stakeholder groups. One nurse participant, P5, said that indicating availability on the chat was vital, so that aides would know when nurses might see a critical communication. But simply indicating online or offline status was not enough – a message sent to an offline nurse might contain critical information, for example, that would have warranted an immediate response if the nurse had been available to read it.

Still, coordinators did see some benefit to chat communication; specifically, they felt the permanence of a chat record would allow aides to look back at previous messages to remind them of directions from their coordinators, thus saving both aides and coordinators time. Centralizing these communications in a digital record would, in our coordinators' views, improve on aides' current practices:

> *"One of the first things we tell aides is to buy a little book from the 99-cent store so they have all their addresses in that book. A lot of aides don't comply with that. When they start working and you tell them something, they just tear off a sheet of paper and write it down, but they don't save it. So by texting like this it seems a little more permanent, I should say."* (P15, coordinator)

Coordinators were also quick to point out that the image of an aide texting on a tablet would not be well received by the client or client's family. We now discuss this issue in detail.

4.7.3 A Trigger for Conflict in Patients' Homes

Recall that during our design sessions, we asked participants about their perspectives on aides working with a dedicated tablet placed either at the home of clients or left with aides to constantly carry with them. Across aide, nurse and coordinator groups, participants agreed that the tablet form factor of our provocation would create problems for aides on-site, in their clients' homes.

Nurses and coordinators alike said they already spent much of their work fielding complaints from clients who felt aides were distracted by their own personal devices, and that the addition of another device, even if ostensibly for work purposes, would make the problem worse. One nurse, P5, said she often performed pop-up visits to client sites in cases where the client's family reported the aide for *"being on their phone all day long."* Coordinators, who by virtue of their roles naturally bore the brunt of client complaints, said they knew aides would *"always be texting"* (P16) if given an additional device. One coordinator said the addition of a work tablet would give aides cover while they continued to text on the clock:

> *"The biggest complaint we get from clients is that aides are always on their phones. 'She don't do nothing, she just sit on her phone all day, texting texting.' So I don't know how I would explain it to the clients. They'd be sneaky, they'd be telling the patients 'Oh, I'm texting my supervisor...' "*
> (P16, coordinator)

Aides, on their part, raised the issue that placing tablets in clients' homes would cause friction in their relationships for a different reason: clients would begin to suspect they were using them for surveillance. As one aide put it:

> *"If this were to get piloted, people would need to be informed, meaning the patients. They can get funky. They can feel like you're spying on them."*
> (P10, aide)

Participants agreed that clarification that aides were using the tablets for task-recording purposes only would need to be provided at the agency level.

However, even in the hypothetical case of agencies successfully reassuring clients about why aides were using the tablet, the tablets themselves would still be a source of conflict. Specifically, aides voiced the concern that patients and family members might steal the tablet if left in the home.

4.7.4 Mismatched Interests in Educational Resources

An important component of our provocation was providing an avenue for aides to access supplemental educational resources on heart failure conditions that they could reference at any time. When we tested this resource with nurses and coordinators, they showed little interest in this part of the provocation. They felt that it was fine to have but not critical for aides to do their jobs.

In stark contrast, the educational resource elicited hearty responses from aides who perceived it as an opportunity to satisfy a long-awaited need to learn more about how they could better do their jobs. In addition, the *behavioral elements* that facilitated decision making made the application more intuitive as aides felt they could better understand when to take actions. For one aide participant (P6), the educational features were "*the best part, because [they] would give me information*". When presented with the educational resources of the provocation, aides were quick to volunteer suggestions beyond what was showed to them. They requested several new pieces of content, with ideas ranging from medication guides to diet information personalized for the client to embedded videos demonstrating how to do cardiopulmonary resuscitation (CPR).

Aides also suggested crowdsourcing the content of the education feature, with one aide, P2, noting that over decades of working as an aide, she had developed an index "*stored in [her] brain*" of side effects of common medications. She asked whether such resources could be made part of the provocation for others' benefit: *"So this would be like a Wikipedia of information that could help people?"*

Compared to nurses and coordinators, our aide participants were quick to ascribe more *trustworthiness* to the information that would come through the provocation than the information they currently utilized on a regular basis. Specifically, aides drew a distinction between what they saw would be provided by the provocation and their current practice of searching the Internet for information.

> *"I don't trust Google. Not everything is accurate. I know this would be designed with accurate information, and help guide me more with it. Knowledge is power."* (P6, aide)

Aides also voiced that such a tool would lend them emotional reassurance as they went about caring for their extremely sick patients. This emotional reassurance, aides said, would be the result of acquiring more knowledge on what they could do for patients with difficult conditions like heart failure:

> *"You'd have something to anchor you and take away your anxiety – 'oh my god, heart failure, what is it, what do I do'. This gives you a base and a foundation."* (P6, aide)

4.7.5 An Additional Workload and Learning Curve

Our provocation was constructed as a medium-fidelity prototype so that it was concrete enough to suggest real functionality, but vague enough to solicit participants' feedback on future possibilities, Our findings show that the fundamental elements of the provocation—a digital application delivered via a tablet—were enough to raise concerns from participants in all stakeholder groups about how aides might receive it if implemented. Participants agreed that, just as they would have to be trained on any new tool, the provocation would require aides to engage in additional learning and practice. But they disagreed on how aides would handle this new learning curve.

Several participants voiced doubts around whether aides would *want* to use the tool. They said they had seen other technologies for aides deployed at their agencies in the past, and were convinced a tool like our provocation would not work in practice because aides would not want to put in the effort of learning to use it. A nurse said:

> *"They don't know how to use the tablet, and they don't want to use it, because most of the time they don't like to use it."* (P11, aide)

At issue was the level of perceived additional work that the provocation created. In a job where they already felt overburdened, aides might be reluctant to engage with the additional responsibilities and expectations that features like blood pressure monitoring and heart failure education would create. This was voiced by not just nurses and coordinators but also aides:

> *"Some people want to go, follow the plan of care, and that's it. Stuff like this would seem like extra work."* (P13, aide)

Participants were also concerned about the English literacy required to engage with a tool like our provocation. All text in our provocation was presented in English, and for the purposes of this study we did not create a Spanish-language version. As one aide who worked in both English and Spanish explained:

> *"Some [aides] don't even know how to write and read in Spanish. How are they gonna be writing and reading in English?"* (P13, aide)

Still, some of the aides felt they had both the ability and desire to expand their capabilities to use a tool like our provocation. Many told us they were already doing additional work outside of agency-mandated courses to learn more about the conditions they worked with each day. As an aide, P14, said: *"We can learn. Why not?"*

4.8 Discussion of the Digital Provocation

Our findings reveal numerous tensions and trade-offs inherent in trying to design to improve equity for home health aides as their role in the U.S. healthcare landscape changes. We also find compelling connections from this work to ongoing discussions within HCI on how to design technology for marginalized workers. Specifically, we contribute lessons from a concrete enactment of Dombrowski et al.'s [68] framework of social justice-oriented interaction design.

First, we note that in engaging with multiple stakeholders and specifically provoking them to react to our vision of their possible futures, we invited a sense of conflict into our polyvocal design process. Coordinators and nurses were quick to point to aides as the source of numerous problems; aides were quick to tell us how difficult it was to get in touch with coordinators and nurses; and all were quick to point out the challenges of dealing with clients and clients families. Constructing a provocation that principally centered on aides also bolstered the turbulent nature of the feedback we received. Here we see a reflection of the *commitment to conflict* that Dombrowski et al. [68] discuss as a prerequisite for designers engaging in social justice-oriented work, as well as an enactment of their strategy of *recognition*. In provoking these conversations, we invited contestation, and in doing so surfaced larger injustices around how aides are valued by the other stakeholders.

Second, we note that as a research team we pursued a shared vision in this project: a healthcare workforce in which aides, specifically, are able to achieve greater parity. We are aware that in doing so, we centered on a specific stance originating from our perceptions of the world. Taking cues from what Dombrowski et al. [68] articulate as a *commitment to reflexivity*, we acknowledge that our visions for aides may not necessarily map onto the visions aides may pursue for themselves.

Nevertheless, our study design was *intended* to provoke participants to contemplate the political; at no point was it construed as the apolitical development of technology against a static efficiency measure or a fixed set of user needs. In taking a stance at all, we align our work with the *commitment to ethics and politics* that Dombrowski et al. [68] outline as a prerequisite for social justice-oriented design practice. Similarly, we claimed responsibility for our design choices and our provocation's underlying assumptions and biases, and we never gave the impression to one stakeholder group that another group had created our provocation. We thus enacted the strategy of *accountability* [68] in the designer-participant relationship.

Our empirical study also provides grounded ways to develop and extend these concepts. We now discuss how our work extends three key strategies within Dombrowski et al.'s [68] framework: 1) the use of methods like ours in designing for *transformation*, as a way to explore how near-future political realities might manifest in marginalized workers' experiences; 2) the need for designers who seek to *enable* marginalized workers to weigh the heightened expectations created by such projects; and 3) the difficulty of designing for *reciprocity, distribution, and accountability between stakeholders* within complex multi-stakeholder environments like ours.

4.8.1 Design to Concretize Transformation

Designing to combat large-scale structural inequities necessitates a shift in focus, from attending to individual needs with user-centered design to attending broadly to oppression-producing contexts. Dombrowski et al. [68] articulate the need for this shift in their strategy of *transformation*, a design practice that not only focuses on "*immediate innovation*" at the level of the individual but also on designing for longer-term collective action. Such an orientation, the authors posit, leaves room for social justice to remain responsive to changing power structures, social norms, and cultural forces.

An approach like this may at first seem distant to the practice of HCI, which has retained a focus on creating and experimenting with digital technologies for people to use—something closer to engineering innovation than to political activism. The balance between these two orientations has been the subject of much literature, notably Bardzell's articulation how feminist interaction design should retain the quality of advocacy [15]. But as designers and builders of technologies, *how can we actualize such orientations and qualities in our work?*

We suggest practices like the design provocation used in this paper as a tactical methodology by which designers might bridge the gap between the political and the material, and address transformation. Our provocation was situated in a near-future shift in healthcare payment models, and concretized a broader transformation in which aides occupy a more valued role in home care.

Within research contexts like ours, home care for heart failure patients, experimentation is limited and heavily regulated due to the life-or-death nature of the work. In these contexts, it is tempting to reduce innovation processes to traditional cycles of building and deploying tools for evaluation against a rigid set of predetermined measures. This is expected, given that people's lives are at stake. Yet, doing so emphasizes measuring outcomes based on incremental improvements on a present reality.

Our method was useful in garnering insights and feedback on both immediate individual needs and longer-term collective goals, as substantiated by our findings. Participants responded to our provocation with feedback proximal to their immediate needs, for example, by describing how the provocation's communication features would overburden nurses and coordinators, or by informing us that placing a tablet in clients' homes would result in conflict. But they also responded with distal feedback highlighting potential broader changes, e.g., how accurate record-keeping could help aides achieve greater parity with their employers, or how aides might crowdsource information on side effects of common medications.

Designing and testing a provocation also gave us a window into the kinds of measures to prioritize in designing a future system that centers on equity for aides as the goal. An intervention that gives aides a way to keep task records, for example, might be evaluated against not just its technical stability for immediate use (”Does it record and store aides' tasks?”) or its user experience (”Can aides use this?”), but also whether it engenders specific equity goals (”Have aides used this to negotiate a raise?”) Such a study might meaningfully address the individual and collective dimensions of the expected transformation while grounding measures in realities sourced from aides themselves. In short, if we supplement near-term measures with equity measures sourced from studies like ours, we might achieve a process of innovation that bolsters broader social transformation.

Further, drawing on Behavioral Science literature to design the provocation made the components more persuasive. For example, by nudging users through different colors of messages displayed (e.g. message in red background showed after *"shortness of breath"* is reported), it is immediately clear the decision that an aide should take. Although it is uncommon in HCI4D to draw on Behavioral Science literature, doing so facilitates the design of technologies that are more persuasive.

4.8.2 Weighing Enablement and Expectations

We now turn to *enablement*, a principle defined in Dombrowski et al.'s [68] framework as a focus on *"fostering human capacity or helping people take advantage of opportunities by creating platforms for participation and self-determination."* We saw this principle enacted in aides' reactions to many of the affordances of our provocation. For some aides, the educational features were seen as a way for aides to fulfill their potential: one aide even responded to these educational features by saying *"knowledge is power."* In addition, aides saw the task recording and chat features as platforms by which they could better control the narrative around their labor.

Despite this, several of the features in our provocation were sometimes seen by participants not as a means for enablement within inequitable labor systems, but rather as a way for additional *expectations* to be placed upon them. For example, the addition of educational instructional videos for CPR may alleviate aides' immediate anxieties around caring for heart failure patients, but, as participants pointed out, aides might feel even more anxious around the newfound expectation to know and carry out CPR.

In this, we see a core tension: *How do we as technology designers balance the goals of enablement with the realities of increasing expectations of marginalized workers?* The answer is not as simple as giving an enabling technology to an employer and encouraging them to make its use optional. It is possible that in deploying a disruptive technology, a subset of marginalized workers might refuse to train themselves to new capacities and be fired as a result. As job loss is not a desired outcome of social justice-oriented technology design, we would define this as a failure of the project.

If we assume that creating the enabling technology is to raise expectations for marginalized workers, then the possible positive outcomes of enablement would need to be worthwhile for workers who did choose to upskill. Yet, our findings show this is not guaranteed. As the reactions to our provocation illuminated, aides may perceive a new technology as *"extra work,"* which would be unwanted if they simply seek stability and predictability in their jobs: *"to go, follow the plan of care, and that's it,"* as one aide described. For these aides, upskilling on new technologies might in fact impose especially impactful burdens on the more marginalized in the labor force. As *improving* the lives of these workers is a core motivation of our work, we would define this, too, as a failure of the project.

Thus we arrive at a final reality: technology enablement projects may only achieve their goals for the subset of users who take and leverage the opportunities created by enablement. To design enablement tools for marginalized workers is to decide what level of upskilling to mandate, to appropriately weigh the promises of enablement with the possible opposing realities of creating additional expectations. New frameworks for how we as technology designers can make these decisions is a compelling area for future work.

4.8.3 Negotiating Reciprocity and Distribution

Designing for equity within complex multi-stakeholder ecosystems like ours also requires foregrounding what Dombrowski et al. [68] call *reciprocity*: a focus on *"relationships and the ways they maybe need to change to become more equitable for all* *stakeholders.*" Design practices with this lens create "*platforms with the potential to change inequitable relations.*" Importantly, designers that seek to engender reciprocity are advised that "*fostering more equitable relationships may not always be seen as mutually beneficial, but can also encourage spaces for open contestation and conflict*."

Closely related to this concept is the strategy of *distribution*, in which designers focus on equitable rebalancing of "*the benefits and burdens of social systems*." Benefits can include wealth, opportunities and access to resources; burdens can include undesirable work, lack of adequate income, and other discomforts. In addition to these concepts, we also find close ties to Dombrowski et al.'s [68] strategy of *accountability*, which emphasizes that "*those who foster or unduly benefit from the oppression of others*" should be held responsible.

Here, we find a tension core to the work: *How do we as technology designers navigate the conflict-ridden space of enabling the most marginalized and ensuring equity for all stakeholders?* How do we do this while reassuring all stakeholders of mutual benefit, so that conflicts over the realities of redistribution do not prevent progress?

Our provocation centered aides as users, but included features impacting nurse and coordinator workflows. Thus, our work presents an empirical view on how open contestation and redistribution between stakeholders can play out in a justice-oriented design practice. By presenting a provocation from an aide's point of view, we elicited perspectives from nurses and coordinators revealing gaps between what they see as aides' current roles and capabilities, and how aides might engage with them under the transformation we concretized.

Predictably, our findings are replete with comments from nurses and coordinators on how aides' current behaviors would cause problems for all stakeholders if they were given a tool like our provocation. For example, coordinators feel better communication would lead to aides sending a deluge of urgent-flagged messages, and nurses point out that they already regularly make house calls to assuage clients who feel aides are constantly texting. In practice, the realities of managing multi-stakeholder projects mean that technology designers themselves must often make trade-offs between emphasizing enablement for marginalized workers and systemic equity for all. Coordinator and aide reactions to the chat feature of our provocation, for instance, seem to demand that we "pick a side."

One way to handle these trade-offs is to create a platform through which stakeholders might themselves engage with the messiness of enacting reciprocity through distribution. However, just as designers of enablement projects must reckon with how their design choices impose implicit expectations upon users, designers of technologies for reciprocity and distribution must also reckon with the idea that designers do not control how these platforms might be used, or whether the eventual outcome of widespread adoption of these platforms is one in which equity has been achieved. For example, it is possible that a tool like the chat feature of our provocation could provide a vector through which employers (e.g., agencies) enact greater surveillance and control over their employees. Another compelling area of future work is developing technology design frameworks around ensuring that when stakeholders have a platform in hand, the resulting human process of redistribution for reciprocity achieves an equitable result, as well as ensuring accountability while leaving room for shifting conceptions of who in the ecosystem are allies, oppressed, and oppressors.

CHAPTER 5

CASE STUDY 3: A LOW-COST TECHNIQUE FOR MITIGATING PARTICIPANT RESPONSE BIAS IN RURAL INDIA

5.1 Introduction

This chapter presents two experiments (one online and the other in the field) that demonstrate a practical and generalizable technique for HCI researchers to mitigate participant response bias and encourage participants to provide critical feedback when evaluating an artifact. HCI researchers and practitioners are increasingly interested in engaging with marginalized communities to design new technologies to have a positive impact on people's lives, including low-income [64, 209], low-literate [165, 208], rural [24, 266, 250], disabled [187, 251], and other communities [54, 101, 264]. One characteristic that these diverse contexts share is that there are frequently large differences between researchers and their participants, such as differences in background, social status, culture, language, education, and technical expertise. Unfortunately, these differences have been shown to substantially impact researchers' efforts to evaluate their new designs or interventions. In particular, usability studies and field evaluations frequently suffer from high levels of participant response bias [56], defined as the extent to which participants provide researchers with feedback or results that will please the researchers or help to achieve the research goals [79, 186]. As a result, many researchers have found it challenging to obtain critical or negative feedback from participants that could help them to improve their designs or interventions [3, 101]. Although participant response bias is present in *all* studies with human participants, its effects have been shown to be significantly amplified in studies involving marginalized communities [56]. Although a growing number of studies acknowledge the potential for participant response bias to impact their results (*e.g.*, [108, 146, 242]), little progress has been made on developing practical tools and techniques that could help HCI researchers to cope with response bias in their studies.

The goal of our research is to fill this gap by contributing a generalizable technique to influence response bias and encourage participants to provide constructive feedback, particularly critical feedback. We conducted a series of controlled experiments that systematically influence participant response bias using the concept of social proof (or informational social influence) from the field of social psychology [61, 217]. Social proof refers to the psychological phenomenon where people assume the actions of others in an attempt to reflect correct behavior in a given situation. In other words, when people are uncertain about what to do, they assume that the people around them, such as experts, celebrities, and friends, have more knowledge about what should be done.

We conducted two controlled experiments: an online experiment with 245 workers recruited through Amazon's Mechanical Turk (MTurk) platform, and a field experiment with 63 low-income, low-literate participants in rural India. Working within an existing HCI project, the *Projecting Health* project in India [137, 140, 253], we asked participants to evaluate a community-created video. In both experiments, participants were randomly assigned to one of the three conditions: positive social proof, negative social proof, and no social proof (*i.e.*, baseline). Prior to watching the video, participants in the positive and negative conditions received social proof through subtle exposure to three positive and negative 'video reviews', respectively, that they perceived to have been provided by other participants 'like them'. Participants in the baseline condition were not exposed to any reviews. We hypothesized that participants in the positive and negative conditions would provide feedback that conformed to the tone of the reviews they encountered. We structured each experiment to examine the effect of social proof on participants' quantitative ratings and qualitative feedback on the artifact being evaluated.

At a high level, our findings show that social proof had a profound effect on participants' evaluations of the artifact in both the online experiment and the field experiment. We found statistically significant differences between the three experimental conditions for *both* the quantitative ratings *and* the qualitative feedback provided by participants. In general, participants in the negative social proof condition gave the video lower ratings and provided a greater amount of critical feedback than participants in the baseline condition. On the other hand, participants in the positive social proof condition gave the video higher ratings and provided a greater amount of positive feedback than participants in the baseline condition. These findings confirm that social proof is an effective way to influence response bias and, in particular, that negative social proof is an effective way to elicit critical feedback from participants, both online and in the field.

Our intervention possesses several key benefits that make it practical for researchers and practitioners to implement. For example, the technique effectively elicits negative feedback even when participants are evaluating a single artifact that is known to be associated with the researcher [56]. It is also a low-cost intervention that does not require any additional equipment beyond the artifact being evaluated. Moreover, the procedure is relatively simple to understand for organizations working in the field and for participants. Finally, by conducting two experiments in different contexts—with MTurk workers online and with low-literate participants in the field—we demonstrate that our intervention could be applied by HCI researchers to a wide range of contexts and domains.

5.2 Related Work

There has been a growing concern within the HCI community about the effects of participant response bias in evaluations of new designs or technological artifacts. A number of studies have discussed the difficulty of eliciting critical or negative feedback from participants, particularly in HCI for Development (HCI4D), where there are often large social and cultural differences between researchers and participants [3, 101, 108, 146]. Brown et al. studied the challenges of conducting HCI trials in *"the wild"* and documented the effects of demand characteristics [186], in which participants adjust their behavior to match the expectations of the researchers. Dell et al. [56] conducted a study in India to quantify the effects of participant response bias, and found that participants were 2.5 times more likely to prefer a technological artifact that they believed to have been developed by the researcher, even when the alternative was identical. In addition, when the researcher was a foreigner who required a translator, the response bias with low-income Indian participants increased to five times. Trewin et al. [242] analyzed participants' subjective Likert-scale responses in accessibility studies, and found that participants in non-anonymous studies gave more positive ratings than those in other studies.

HCI researchers have suggested a variety of approaches to try and reduce participant response bias. Brown et al. [29] suggested postponing the evaluation of technologies altogether until the technologies can be better understood by users. Chavan [37] encouraged participants to submit critical feedback by situating user studies within dramatic storylines. Molapo et al. [173] recommended role playing and skits to motivate frontline workers to share their opinions. Other researchers have explored reducing response bias by dissociating themselves from designs or artifacts [189, 254], limiting direct contact with participants [81, 252], or spending more time with participants in the field in the hope that they would be comfortable enough to provide critical feedback [76]. However, for the most part, the impact of these approaches on reducing response bias has not been systematically quantified.

Our study uses the concept of social proof from the field of social psychology to influence response bias and encourage participants to provide constructive, critical feedback to researchers. Social proof [217] refers to the psychological phenomenon of assuming the actions of others in an attempt to reflect correct behavior. Also known as informational social influence, social proof occurs when people experience uncertainty about what decision they should make, assume that the people around them possess more (or better) information, and accept information gleaned from other people's behavior as evidence about reality [61, 63]. Examples of social influence include presuming that the food at a restaurant is good because the queue is long, endorsing a political candidate because everyone else approves of the person, or giving a product excellent reviews because an expert or celebrity positively reviewed the same product. The effects of social proof have also been shown to differ across countries and cultures [40]. For example, prior research has demonstrated that people living in collectivist cultures (such as India) tend to conform to social proof more often than those in individualist cultures [27].

There is a growing interest within the HCI community in understanding and applying the concept of social proof to a range of application domains, such as interpreting graphical information and visualizations [105], influencing user opinions in recommendation systems [44], prompting people to explore and adopt better security habits [48, 49], and affecting people's intention to adopt privacy behaviors [168]. Several scholars have also studied social proof, or the broader concept of social influence, in the context of online platforms. For example, Bond et al. [26] found that showing people that their Facebook friends have voted increased voter turnout. Burke et al. [34] showed that social learning played an important role in influencing how novice Facebook users interact with the platform. Kramer [134] found that people were more likely to share emotional content that matched the content shared by their friends. Malu et al. [159] used social influence to encourage people to contribute personal content to an online community. Finally, Wu and Huberman [265] examined social influence in the context of online opinions, news, and product reviews, and found that awareness of others' opinions leads to increasingly extreme views. Our paper extends this body of work by conducting controlled experiments that measure the impact of social proof in the evaluation of an HCI artifact. To the best of our knowledge, ours is the first paper to apply the concept of social proof to influence response bias in HCI. We are also the first to study the effects of social proof with low-literate populations in resource-constrained settings.

5.3 Intervention Design

We situated our study in the context of *Projecting Health*, an existing community-driven social and behavior change intervention to improve maternal and neonatal health in rural India [137, 140, 253]. *Projecting Health* empowers community-based organizations to produce videos that feature local people discussing key health messages in a local dialect. Accredited social health activists (ASHAs) share the videos in group sessions with women via portable projectors. The project is currently operating in over 125 villages in Uttar Pradesh with 170 mother groups. Thus far, 80 videos have reached an estimated 100,000 people through 12,000 screenings.

A critical component of *Projecting Health* is to obtain feedback from stakeholders to ensure that videos are suitable for dissemination in rural areas. During the initial phase of the project, several participants attended video disseminations out of the curiosity to watch videos featuring people 'like them', and also because of the novelty of accessing health information via videos. Since these effects lead only to short-term engagement, the *Projecting Health* staff has aimed to design improved videos that low-income, low-literate women find engaging, interesting, informative, and entertaining. However, the staff has reported great difficulties in obtaining any critical feedback from rural women because of high levels of participant response bias. Often they receive positive feedback, or feedback that lack details. During an informal conversation in the field, the program manager of *Projecting Health* described:

> *"The biggest challenge [in Projecting Health] is to improve the quality of the videos. If a video is of good quality, useful, and entertaining, people will automatically watch it again and share it with others. However, it is almost impossible to get constructive feedback in rural areas. They [people in rural areas] always say the video is very nice and there is no need of improvement."*

The goal of our research is to contribute techniques for influencing response bias and encouraging participants to provide constructive, critical feedback. A key design consideration is to ensure that the intervention is easy to administer and generalizable to a variety of settings.

To this end, we began our research by reviewing Behavioral Science literature for ways to persuade users to share critical and honest feedback. In addition, we consulted two behavioral experts for suggestions on addressing response bias. To narrow down our list of potential behavioral methods, we eliminated techniques that were challenging to implement and difficult to scale with technology, and arrived at *priming* [226] and *social proof* [217] as promising techniques. Next, we tested each of these concepts in a controlled online experiment, which revealed that *social proof* effectively elicited critical feedback from users but not *priming*. We now focus the rest of this chapter on the design of the social proof intervention.

To test the social proof intervention, we conducted a between-subjects study where participants were randomly assigned to one of the three conditions: positive social proof, negative social proof, and no social proof (*i.e.*, baseline). Participants in the positive and negative conditions were subtly exposed to a set of positive and negative video reviews, respectively. In reality, we authored the reviews in collaboration with the *Projecting Health* team, and experimentally controlled their content and tone to provide participants with either *positive* or *negative* social proof. For example, a review that we created to provide participants with positive social proof is: *"It is very important for people to learn this information. The video content is great! The health messages are very easy to understand."* By contrast, an example of a review that we created to provide participants with negative social proof is: *"Nobody can understand the content of this video. The message is not clear. This will never help anyone."* We hoped that showing participants these 'reviews', which they perceived to have been given by other participants 'like them', would encourage them to provide their own feedback on the video. In particular, we hypothesized that if participants perceived that other people had contributed negative feedback, they may feel comfortable to critique the artifact being evaluated.

After participants received positive, negative, or no social proof, they watched a three-minute *Projecting Health* video about safe drinking water. The video featured a discussion between an ASHA, two representatives of a village-level committee, and a local doctor on how to keep ground water clean. The *Projecting Health* staff recommended this video since it had both strengths (*e.g.*, important topic and new knowledge for most people) and weaknesses (*e.g.*, unskilled actors and uninteresting storyline). After watching the video, participants completed a survey in which they provided quantitative ratings of the video along with unstructured qualitative feedback.

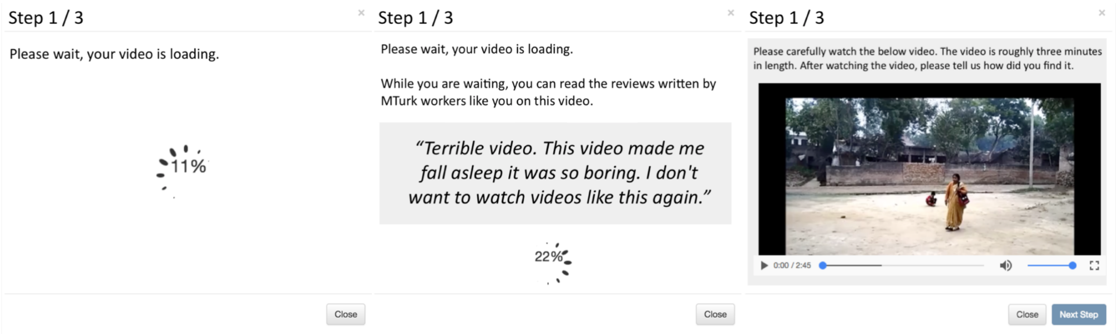
We conducted two experiments to evaluate the impact of our social proof intervention with participants in different contexts: (1) an online study with MTurk workers, and (2) a field study with low-income women in rural India. Each experiment focused on answering the following research questions:

RQ1: How does social proof impact participants' quantitative ratings of an intervention?

Many HCI studies evaluate new designs, products, or interventions by asking participants to rate their subjective experiences or opinions on the intervention using quantitative instruments such as a Likert scale [149]. We hypothesized that participants' quantitative ratings of a *Projecting Health* video would be influenced by the kinds of reviews that they saw before watching the video. For example, participants who were exposed to negative video reviews would submit more negative ratings than those who were exposed to positive reviews.

RQ2: How does social proof impact the qualitative feedback provided by participants?

We hypothesized that participants would be influenced to provide qualitative feedback of a tone similar to the reviews that they saw before watching the video. For example, participants who saw negative reviews would provide more



Step 1 / 3

Please wait, your video is loading.

Step 1 / 3

Please wait, your video is loading.

While you are waiting, you can read the reviews written by MTurk workers like you on this video.

“Terrible video. This video made me fall asleep it was so boring. I don't want to watch videos like this again.”

Step 1 / 3

Please carefully watch the below video. The video is roughly three minutes in length. After watching the video, please tell us how did you find it.

Figure 5.1: Screenshots from the MTurk experiment (shown in English for readability, although the experiment was in Hindi). On the left, a video is loading without any reviews (baseline). In the middle, a video is loading with negative reviews. In the third image, a video is playing after it has loaded.

negative qualitative feedback than those who saw positive reviews.

5.4 Experiment 1: Study on MTurk

Our first experiment analyzed the impact of social proof in an experiment conducted with 245 participants recruited through MTurk—an online crowdsourcing marketplace where workers complete tasks such as categorization, translation, and surveys in exchange for small monetary payments [1]. An increasing number of HCI studies recruit MTurk workers as participants [127, 130, 163] since MTurk makes it easy to recruit large numbers of geographically distributed populations at a relatively low cost. Since the prevalence of HCI studies conducted on MTurk is rapidly increasing, we examined how social proof might impact the evaluation of an HCI artifact by MTurk workers.

5.4.1 Authoring and Validating Reviews

In collaboration with the *Projecting Health* staff, we authored thirty positive and thirty negative reviews in Hindi that commented on the video’s production quality, content, acting, storyline, duration, and entertainment value. The positive and negative reviews were similar in length and attributes being evaluated. The average length of reviews was 26 words (*SD* = 6 words). To ensure that the reviews were perceived as positive or negative, we recruited 125 MTurk workers from India. Each worker was randomly assigned ten reviews to read and rate on a five-point Likert scale from very negative to very positive. Since the reviews were in Hindi, we restricted participation to MTurk workers who could understand Hindi by providing the instructions and prompts in Hindi.

Workers who rated the reviews were 32 years old, on average. Eighty-eight workers were male, 34 were female, and three did not indicate their gender. One worker had completed secondary school, three had completed high school, 76 had finished a bachelor's degree, and 45 had finished a master's degree. The positive reviews received an average rating of 4.6 (*SD* = 0.23) while the negative reviews received an average rating of 1.7 (*SD* = 0.31). For the final experiment, we selected the ten highest rated and ten lowest rated reviews.

5.4.2 Procedure

Since the *Projecting Health* video as well as the reviews were in Hindi, we restricted participation to MTurk workers who were located in India, and were comfortable reading and understanding Hindi. To participate in our study, MTurk workers needed to answer a basic arithmetic question (*i.e.*, what is ten plus seven) displayed in Hindi. Workers who provided the correct response were directed to an external webpage that contained the study instructions and prompts in Hindi.

Each consenting MTurk worker was randomly assigned to one of the three experimental conditions: positive social proof, no social proof (*i.e.*, baseline), or negative social proof. We balanced these three groups on participants' income, age, and education. Before showing participants the *Projecting Health* video, we purposefully introduced a thirty-second delay that we told participants was due to the video loading. In the baseline condition, participants simply saw a progress bar that took thirty seconds to reach 100% (see first image on the left in Figure 5.1). In the positive and negative conditions, we used the delay to show participants three randomly selected reviews, each for ten seconds (see middle image in Figure 5.1). After the thirty-second period was over, participants in all three conditions watched the video and provided their feedback (see third image on the right in Figure 5.1). We requested participants to rate the video using a five-point Likert scale on four parameters: how much they liked or disliked the video (*likeability*), how useful the video was (*usefulness*), how entertaining the video was (*entertainment value*), and how much the video could be improved (*scope of improvement*). We also asked participants to share their subjective feedback on the video. To filter participants that might not have paid attention to the video, we asked a simple validation question about the subject matter of the video. We also collected participants' demographic information. The experiment lasted for around ten minutes and participants received USD 1 for their participation.

5.4.3 Participant Demographics

We recruited 245 MTurk workers for our experiment, with 84, 73, and 88 participants in the positive, baseline, and negative conditions, respectively. Since seven participants in the positive condition, and ten participants each in the baseline and negative conditions answered the validation question incorrectly, we removed their responses from our analysis. Table 5.1 shows the demographic characteristics for the MTurk participants who answered the validation question correctly. Participants came from sixty cities in India. All participants had access to a mobile phone and 45% of them shared their phone with family members. Almost 90% of them watched videos regularly and 97% had access to mobile Internet.

5.4.4 Data Analysis

We conducted a single-factor, between-subjects experiment with three levels. The single factor was *type of social proof* with the levels *positive*, *baseline*, and *negative*. We used non-parametric Kruskal-Wallis tests [136] to analyze differences in participants' Likert-scale ratings on likeability, usefulness, entertainment value, and scope of improvement. Post-hoc pairwise comparisons were conducted using Dunn's tests [73] with Bonferroni correction [72] for multiple comparisons.

We analyzed participants' qualitative feedback along several dimensions, including the number of participants who submitted feedback, the length of the feedback, the tone of the feedback, and whether participants provided *substantive* feedback. We defined feedback as *substantive* if participants provided concrete details on what they liked or disliked about the video or suggested specific points for improving it. To analyze the qualitative feedback, we recruited three Hindi speakers (1 male and 2 female) who read each review independently in a random order, and classified the tone of the feedback as positive, negative, or mixed, and noted whether the feedback was substantive. The reviewers were blinded to the experimental conditions. We used majority voting to break ties, and analyzed differences between the experimental conditions using Pearson's Chi-squared tests [190] or Fisher's exact test.

Condition	No of workers	Male (%)	Age (years)	Education (years)	Family Income (USD/year)
Baseline	63	68	31	15.6	1191
Positive	77	71	32	15.4	1116
Negative	78	75	33	15.6	1100

Table 5.1: Demographic characteristics of MTurk participants.

5.4.5 Results of MTurk Experiment

RQ1: Impact on Participants' Quantitative Ratings

Our first research question focuses on understanding the impact of the social proof intervention on participants' quantitative ratings of the video. Table 5.2 shows that participants in the positive condition rated the video highest on *likeability*, *usefulness*, and *entertainment value*. In contrast, participants in the negative condition rated the video lowest on *likeability*, *usefulness*, and *entertainment value*. Participants in the negative condition found greater scope for improving the video than participants in the other conditions. Results of Kruskal-Wallis tests indicated that these differences were significant for all four parameters: *likeability* (p ¡ .001), *usefulness* ($p = .001$), *entertainment value* (p ¡ .001), and *scope of improvement* (p ¡ .001). Post-hoc pairwise comparisons between experimental conditions indicated significant differences between the positive and negative conditions, and the negative and baseline conditions, for all parameters (see Table 5.3). These findings suggest that negative social proof effectively decreased participants' quantitative ratings of the video.

RQ2: Impact on Participants' Qualitative Feedback

Our second research question focuses on understanding the impact of social proof on the qualitative feedback provided by participants. We found that a

Condition	**Like-ability**	**Useful-ness**	**Entertainment value**	**Scope of improvement**
Baseline	3.7	3.8	3.4	3.1
Positive	4.1	3.9	3.6	2.9
Negative	3.2	3.2	2.8	3.7

Table 5.2: Average Likert-scale ratings of the video by participants in the MTurk experiment.

Condition	Baseline	Negative
Positive	$L^{\S}$	$L^{\S}$ $U^{\dagger}$ $E^{\S}$ $S^{\S}$
Baseline		$L^{\S}$ U^{*} E^{*} $S^{\dagger}$

Table 5.3: Pairwise comparison of experiment conditions on (L)ikeability, (U)sefulness, (E)ntertainment value, and (S)cope of Improvement (* is $p < .05$, † is $p < .01$ and § is $p < .001$).

greater percentage of participants provided feedback in the positive (69%) and negative (76%) conditions than in the baseline condition (63%). In addition, the average length of feedback submitted by participants in the positive condition (20 words) and negative condition (19 words) was greater than the baseline condition (17 words). This may indicate that participants who were exposed to other reviews wrote longer feedback since they wanted to conform to other workers who submitted the subjective feedback. However, these differences were not statistically significant for either the number of participants who gave feedback or the length of the feedback.

Table 5.4 shows the classification of participants' qualitative feedback as positive, negative, or mixed (*i.e.*, it contained both positive and negative elements). An example of a participant's negative feedback is, *"The conversation was very unnatural. The flow of ideas can be improved. Dialogue delivery can be improved."* By contrast, an example of mixed feedback is:

> *"This video contained good information most of which I was unaware of. It*

> *was useful for me, but the video could be improved using graphics and other video enhancing ways. The current video is plain and monotonous. "*

Participants in the positive condition submitted more positive and mixed comments, and fewer negative comments, than those in the baseline condition. In contrast, participants in the negative condition submitted more negative and mixed comments, and fewer positive comments, than those in the baseline condition. These differences were significant ($\chi^2(4, N = 152) = 23.2$, $p < .0001$), which indicates that negative social proof led participants to submit more negative qualitative feedback, and vice versa for the positive condition.

The qualitative feedback provided by participants was also classified as either being substantive (*i.e.*, containing concrete suggestions or discussion) or not. An example of feedback that was *not* substantive is, *"This is a good video,"* while an example of a substantive feedback is:

> *"Very nice video that gives us a very important message. Disease is spreading in village due to polluted water. Hand pumps should be very deep and we should try to keep the surrounding area very neat and clean."*

Table 5.4 shows that 74% of participants in the positive condition and 85% of participants in the negative condition provided feedback that was judged as substantive, compared to 68% of participants in the baseline condition. These differences were not statistically significant though. Analysis of the negative and mixed feedback indicated that participants provided several suggestions, such as improving the acting (N=48), creating interesting storyline (N=24), enhancing entertainment value (N=16), and adding graphics and examples (N=8), among others. Analysis of the comments that contained positive and mixed feedback indicated that 81 participants found the video useful and informative, seven

Condition	Total # comments	Positive feedback	Mixed feedback	Negative feedback	Substantive feedback
Baseline	40	65%	25%	10%	68%
Positive	53	68%	24%	8%	74%
Negative	59	29%	41%	30%	85%

Table 5.4: Classification of the feedback provided by participants in the MTurk experiment.

liked the location where it was shot, and five appreciated the acting skills of people with rural background featured in the video.

In summary, our experiment with MTurk workers demonstrated that social proof influenced participants' quantitative ratings and improved their qualitative feedback. Participants who were exposed to positive reviews perceived the video more positively, provided positive ratings, and supported their ratings with substantive positive comments. Similarly, participants who were exposed to negative video reviews were more critical of the video, submitted lower ratings, and wrote substantive negative and mixed feedback critiquing the video.

5.5 Experiment 2: Field Study in Rural India

Our second experiment examined how social proof might impact a field study in which a researcher conducts a face-to-face evaluation of an HCI intervention with participants. In particular, prior research has shown that evaluations of HCI artifacts with participants in HCI4D contexts may suffer from high levels of participant response bias [56]. Our work directly engages with these contexts through an *in situ* experiment with low-income, low-literate women in rural India.

Figure 5.2: The three stages of the field experiment in India. On the left is a waiting area for participants. In the middle, two confederates are social proofing a participant. In the third image, a participant is watching the video projected on wall in a research room.

The second confederate is reading the feedback forms. In front of the participant, the first confederate asks the second:
First confederate: *"Sister, what are you reading?"*
Second confederate: *"Brother, yesterday we went to* [neighboring village] *where the researcher showed the three-minute Projecting Health video. He asked women for their feedback on the video and noted it down. I was just reading the feedback women gave to him. See, this women told him* [the confederate randomly selects a form and reads the feedback]"
First confederate: *"Hmm. What did others say?"*
Second confederate: *"Several women gave feedback. See* [points at another page], *this woman said* [reads a second review]"
First confederate: *"Hmm . . . Okay . . . "* [appreciating nod]
Second confederate: *"Yes brother, another sister told* [confederate selects and reads a third review]"
Second confederate turns to the participant.
Second confederate: *"Such detailed feedback is very important to improve the project. You should give your feedback without any hesitation like these women in the neighboring village did. He will also ask you information to fill this form. You should tell him what you like and what you don't like freely."*

Table 5.5: Translation of script used by the confederates to social proof participants.

5.5.1 Authoring and Validating Reviews

Conducting our experiment within the context of an ongoing HCI project introduced a number of considerations. In particular, the *Projecting Health* staff requested that we create reviews for the experiment that do not critique key aspects of *Projecting Health*'s design such as the use of local dialect and actors. Thus, we authored a new set of 15 positive and 15 negative reviews that focused only on other video attributes like production quality, content, storyline, duration, and entertainment value. The positive and negative reviews were similar in length as well as the attributes being evaluated. The average length of the reviews was 30 words (SD = 6 words).

To ensure that the reviews were successfully perceived as positive and negative, we recruited three *Projecting Health* staff to read each review in a random order and rate it on a five-point Likert scale from very negative to very positive. The staff members (1 male and 2 females) were native Hindi speakers and had completed master's degrees. They had been associated with *Projecting Health* since its inception and had a deep understanding of the rural communities it serves. The positive reviews received an average score of 4.5 (SD = 0.33) and the negative reviews received an average score of 1.6 (SD = 0.4). For the field experiment, we selected the ten highest rated and ten lowest rated reviews.

5.5.2 Procedure

With the support of NYST, a grassroots organization implementing *Projecting Health* in rural Uttar Pradesh, we recruited 63 low-income, low-literate women to participate in the field experiment. Typically, the *Projecting Health* staff show videos to community members using a portable projector and request feedback from them to improve the videos. We designed our procedure to mimic this existing feedback collection routine.

To avoid contamination and confusion, we wanted to ensure that participants in one condition are unaware of the activities assigned to participants in other conditions. This was easy to execute in the MTurk experiment since the participants were geographically distributed and used their own personal computers. However, in rural India, which has a highly collectivist culture, assigning participants to different conditions without contamination and confusion was challenging, especially since the field staff reported that women often come together in a group to watch the videos. We also could not share the purpose of our research experiment with participants beforehand since doing so might have influenced the study outcome. Thus, to avoid any contamination and confusion among participants, we along with the field staff identified three villages that were comparable to each other in terms of socioeconomic status, education, population size and distribution, and availability of resources such as health centers and schools. We then randomly assigned each village to either positive social proof, no social proof, or negative social proof condition, with all participants in the village assigned to the same condition.

In each village, the local ASHA asked women to attend the screening of a new *Projecting Health* video. Once we had a quorum, a local staff member told women that a researcher will show a three-minute *Projecting Health* video one by one to each participant, and ask questions to understand the strengths and weaknesses of the video. Participation in the experiment was voluntary. We asked consenting participants to wait in a specific area for their turn (see first image on the left in Figure 5.2). We set up a portable projector and speakers in a room for the researcher (male, 30 years, native Hindi speaker) to show the video and ask questions (see middle image in Figure 5.2). To ensure that participants were *subtly* exposed to feedback that they perceived to be from people 'like them', we set up a staging area where we asked the participant who would be next to visit the researcher to wait for her turn (see third image on the right in Figure 5.2). In the staging area, two staff members, acting as confederates, were tasked to social proof the participant by acting out the skit described below.

The first confederate (male, 26 years) invited the participant to the staging area and asked her to sit next to the second confederate (female, 32 years) while she waited for her turn to interact with the researcher. The second confederate pretended reading a stack of feedback questionnaires that, in reality, contained the video reviews we had authored. The first confederate then casually inquired what the second confederate was reading. The second confederate replied that she was reading the feedback received from women in a neighboring village where the same activity was conducted yesterday. She then randomly selected three reviews and shared them with the first confederate and the participant. After casually reading the three reviews, she asked the waiting participant to also share her honest feedback with the researcher. Table 5.5 shows the script used for the experiment. The skit lasted less than three minutes. We conducted ten rehearsals with the confederates to ensure that the skit appeared natural and finished on time. We decided against sharing the reviews with all participants as a group to ensure that each participant experienced approximately the same amount of delay between exposure to the reviews and interacting with the researcher. Only participants in the positive and negative conditions were exposed to the reviews. Participants in the baseline condition just waited for their turn while sitting next to the confederate.

After the researcher finished the study with the previous participant, the confederates sent the waiting participant to the researcher's room. The researcher showed the video to the participant, and then requested her to rate, on a five-point Likert scale, how much she liked the video, and what is the scope of improvement in the video. We only requested ratings on *likeability* and *scope of improvement* because the *Projecting Health* staff considered these two questions to be critical for their feedback process, and because they suggested that we limit the number of questions to reduce the time required to participate in the study as well as the possibility of confusing participants. In addition to quantitative ratings, the researcher also recorded qualitative feedback and demographic details. The entire interaction lasted around ten minutes.

5.5.3 Participant Demographics

Overall, 63 low-income, low-literate rural women participated in the field experiment, with 20 in the positive condition, 22 in the baseline condition, and 21 in the negative condition. The majority (78%) owned a mobile phone while the rest used phones of family members. About 25% of participants reported watching videos on their phone, and only four had Internet access. Although 80% of the participants had previously watched a *Projecting Health* video, none of them had seen the video we used in the experiment. About 75% of the participants were homemakers, and the rest were farmers (N=8), laborers (N=3), domestic helpers (N=2), a cook (N=1), tailor (N=1), and teacher (N=1). Table 5.6 shows that participants possessed low levels of education and family income.

5.5.4 Data Analysis

We used the same statistical tests and procedures as the online experiment, including non-parametric Kruskal-Wallis tests to analyze differences in Likert-

Condition	No of people	Age (years)	Family size	Education (years)	Family Income (USD/year)
Baseline	22	36	5.6	5.9	96
Positive	20	31	6.2	5.4	104
Negative	21	29	7.4	5.7	119

Table 5.6: Demographics of participants in the field experiment.

scale ratings, and Dunn's tests with Boneferroni correction for post-hoc pairwise comparisons. Qualitative feedback provided by participants were classified as containing positive, negative, mixed, and substantive feedback, and differences between conditions were analyzed using Pearson's Chi-squared tests or Fisher's exact test based on the values obtained in different conditions.

5.5.5 Results of the Field Experiment

RQ1: Impact on Participants' Quantitative Ratings

Table 5.7 shows that participants in the positive condition rated the video higher on *likeability* and lower on *scope of improvement* than the other two conditions. Conversely, participants in the negative condition rated the video lower on *likeability* and higher on *scope of improvement* than the other two conditions. A Kruskal-Wallis test also indicated significant differences in three conditions on *likeability* (H(2) = 22.5, p ¡ .0001) and *scope of improvement* (H(2) = 7.6, p = .02). Post hoc tests with a Bonferroni correction indicated a significant difference (p ¡ .001) in *likeability* for pairwise comparisons of all three conditions, and a significant difference (p = .02) in *scope of improvement* between the positive and negative conditions. These findings show that social proof effectively impacted participants' quantitative ratings of the video.

Condition	Likeability	Scope of improvement
Baseline	4.3	1.8
Positive	4.6	1.3
Negative	3.1	2.3

Table 5.7: Average Likert scale ratings for *likeability* and *scope of improvement* by participants in the field experiment.

Comparing Quantitative Ratings Online vs. in the Field

We compared the ratings obtained in the field with those obtained in the online experiment. We found that the average *likeability* rating in the baseline condition of the field experiment was 4.3, which was significantly higher than the equivalent rating of 3.7 in the online experiment (H(1) = 5.1, p = .02). Moreover, the average score for *scope of improvement* in the baseline condition of the field experiment was 1.8, significantly lower than the equivalent score of 3.1 in the MTurk experiment (H(1) = 18.3, p ¡ .0001). This suggests that either participants in the field genuinely liked the video more than the participants on MTurk, or that the response bias was much higher in a face-to-face field study with low-income, low-literate participants.

RQ2: Impact on Participants' Qualitative Feedback

Since we asked participants to provide qualitative feedback face-to-face, every participant provided at least some feedback, albeit with varying length and quality. Although some participants just said one word (*e.g., "good"*), many others gave detailed responses (*e.g.,* the longest feedback had 91 words). The average length of feedback was greater in the negative condition (45 words) than the positive condition (32 words) and the baseline condition (16 words). This difference was significant (H(2) = 25.4, p ¡ .001), and post-hoc pairwise comparisons with Bonferroni correction showed significant differences between all conditions (all *p* ¡ .001). This suggests that social proof, particularly negative social proof, successfully encouraged participants to provide more qualitative feedback.

Table 5.8 summarizes the classification of the content and tone of participants' qualitative feedback. Participants in the negative condition provided more mixed and negative feedback than those in the baseline and positive conditions. These differences were significant (*p* ¡ .0001, Fisher's exact test), with post-hoc pairwise comparisons yielding significant differences between the positive and negative conditions (*p* = .002), and the baseline and negative conditions (*p* ¡ .001). These findings indicate that negative social proof successfully encouraged participants to provide critical feedback on the video.

With respect to sharing concrete ideas for improving the video, Table 5.8 shows that more participants in the positive and negative conditions provided substantive feedback than participants in the baseline condition. These differences were statistically significant (*p* = .002, Fisher's exact test), with post-hoc pairwise comparisons revealing significant differences between the positive and baseline conditions (*p* = .01), and the baseline and negative conditions (*p* = .01). These findings show that exposure to reviews prompted participants to provide substantive suggestions for how to improve the video.

Our analysis of positive and mixed comments revealed that a majority of participants (N=43) found the video informative. Eight participants appreciated the local actors and efforts they *"put in to provide information while working tirelessly."* Other participants appreciated the production quality of the video and its entertainment value. A 28-year-old low-literate homemaker who was assigned to the positive condition suggested:

> *"I liked that information about diseases was given. The video taught us*

Condition	Positive feedback	Mixed feedback	Negative feedback	Substantive feedback
Baseline	21	1	0	13
Positive	17	3	0	19
Negative	7	11	3	20

Table 5.8: Classification of the content and tone of participant feedback in the field experiment.

> *that we should not drink unsafe water and consume only clean water. I learned that we should use borewells that are deeper. I liked this information. You should also add songs. You should also add information about what precautions to take with tap water."*

Our analysis of negative and mixed reviews revealed that participants' comments contained actionable suggestions for how to make the video better, with many comments suggesting that *"of course, the video can always be improved."* Nine participants found the key health messages to be overwhelming since they felt that the video was *"rushed"* because of its short duration. Five participants suggested adding demonstrations (*i.e.*, acting things out instead of talking) to make the video more appealing. A 31-year-old low-literate homemaker who was assigned to the negative condition stated:

> *"A lot of information was not given in the video. The information was shared quickly in three minutes, making it difficult to remember. If they demo actions, show clean places, then it would be easier for us to understand."*

Another five participants suggested adding information about other related health subjects. Four participants recommended adding songs, dances, comedy, and photos of children to make the video more entertaining. Another four participants complained about the production quality, two did not like the acting, and one suggested using a more refined Hindi dialect. In summary, our analyses show that the social proof intervention effectively encouraged participants to submit greater amounts of qualitative feedback that contained useful and actionable suggestions for how to improve the video.

Although a few participants echoed the reviews they saw during the social proof exercise, most provided valuable feedback, including detailed suggestions for improving particular video attributes, new topics for future videos, detailed information on high-level themes they heard via social proof reviews, and concrete suggestions to improve the video. For example, although none of the social proof reviews mentioned demonstrations, several participants noted how demonstrations could improve their learning of the subject matter. Similarly, several participants recommended creating videos on new topics like nutritious food and waterborne diseases. Such suggestions were absent from the reviews we used for the social proof exercise. Although some participants did give feedback based on themes they heard via social proofing, they often shared specific details that expanded on these themes. For example, a participant who was exposed to a positive review containing *"I liked the information shared in the video"*, explicitly mentioned in her feedback that she *"liked the information that the hand-pump should not be broken and the house area should not be littered."* Finally, the social proof intervention gave agency to participants to make concrete suggestions that were absent from the social proof reviews, such as including songs or dances to make the video more engaging.

Comparing Qualitative Feedback Online vs. in the Field

We compared the qualitative feedback received in the field with that obtained in the MTurk experiment. Our findings show that the average length of the feedback received in the baseline conditions for both experiments was comparable: 16 words (*SD*=10) in the field experiment vs. 17 words (*SD*=11) in the online experiment. However, the length of the feedback received was significantly higher for the positive condition in the field (32 words, *SD*=14) vs. online (20 words, *SD*=9), (H(1) = 12.2, *p* ¡ .001)). We found a similar trend for the negative condition: 45 words (*SD*=25) in the field vs. 19 words (*SD*=10) online, (H(1) = 23.5, *p* ¡ .0001)). These differences could be due to the obligation the field participants may have felt to provide more feedback since they were face-to-face with the researcher and because others 'like them' were also providing feedback. Similar to the field participants, the feedback from the MTurk participants highlighted new topics and issues not present in the social proof reviews they saw.

5.6 Discussion

5.6.1 Reflection

The goal of our research is to contribute a technique for influencing response bias and encouraging participants to provide constructive, critical feedback to researchers. We created an intervention that introduces social proof by subtly exposing participants to different kinds of feedback that they perceived to have been provided by other participants 'like them'. We evaluated the impact of our intervention through two controlled experiments: an online experiment with 245 MTurk workers and a field experiment with 63 low-income, low-literate women in India. At a high level, our findings show that social proof had a significant effect on participants' evaluations of an HCI artifact, both online and in the field. We found statistically significant differences between positive social proof, negative social proof, and no social proof conditions in *both* the quantitative ratings *and* the qualitative feedback provided by participants. Participants who were negatively social proofed provided lower ratings along with substantive, critical comments, while participants who were positively social proofed provided higher ratings and substantive, positive comments. These findings confirm that social proof is an effective way to influence response bias and, in particular, that negative social proof is an effective way to elicit critical feedback from participants, both online and in the field.

The feedback enabled the *Projecting Health* staff to understand the strengths and weaknesses of different video attributes, such as the production style, choice of accent, and informational content. In addition, the staff learned about specific topics for follow-up videos that would be of interest to their target population. Most importantly, the intervention made it easier for participating women to feel comfortable providing constructive, specific, actionable, and critical feedback. Such feedback had previously been very challenging for the *Projecting Health* staff to obtain, probably because the women were thankful for their efforts and did not want to hurt their feelings. The *Projecting Health* staff found the feedback they received very valuable. For example, based on the numerous suggestions from participants to make the videos more engaging, ten staff members of the grassroots organization took part in a three-day video production training in November 2017.

There were two main differences in our findings between the online and the field experiments. First, the baseline ratings in the field experiment were significantly higher for *likeability* and lower for *scope of improvement* than the corresponding ratings in the MTurk experiment. Since the *Projecting Health* video was exclusively designed for low-income, low-literate women, these differences in ratings could either because participants in the field genuinely liked the video more than participants online, or could be attributed to higher levels of response bias in the field. Second, although the amount of qualitative feedback provided by participants in the baseline condition was comparable between the field experiment and the online experiment, participants in the positive and negative conditions in the field provided significantly more feedback than the corresponding conditions online. This suggests that the social proof intervention, combined with the face-to-face nature of the interaction (*e.g.,* asking questions in-person), encouraged participants to provide more qualitative feedback.

Our social proof intervention has a number of key benefits that make it practical for researchers and practitioners to implement. One of our aims was to create an intervention that is generalizable and reproducible. We demonstrated that our intervention can be applied in two distinct contexts—an online experiment and a field study with low-literate participants in resource-constrained settings. In both experiments, we used the same experimental procedure with minor variations and received similar results that prove the efficacy of our social proof intervention. Compared to other techniques that aim to reduce response bias (*e.g.,* randomized response [259] and unmatched count [200]), our intervention is low-cost, practical, easy to understand for organizations and participants, reproducible in different contexts (as we demonstrated), effective for both quantitative and qualitative feedback, and elicits critical feedback even when participants are evaluating a single artifact that is known to be associated with the researcher [56]. Taken together, these benefits suggest that, with a small amount of adaptation (described below) the intervention could be used by HCI researchers in a wide range of contexts and domains.

5.6.2 Challenges and Design Recommendations

We now discuss the challenges we faced in executing our experiments, and recommendations for researchers interested in using our intervention. Although the intervention is designed to be an add-on to any artifact being evaluated, its efficacy is dependent on how *subtly* participants are exposed to social proof. On MTurk, we introduced a fake delay in loading the video and used that time to show video reviews. In the field, we created a skit that exposed participants to feedback that they perceived to have been provided by women in neighboring villages. Researchers in other contexts will need to find new ways to subtly expose participants to social proof.

We encountered several practical challenges in the field. For example, it was difficult to find sufficient space to conduct the experiment. We needed three physical spaces (waiting area, staging area, and researcher's room) to avoid contamination and confusion among the participants. We coped with the space challenge by using verandahs, porches, and lawns as waiting or staging areas. We were also concerned about participants returning to the waiting area after interacting with the researcher and sharing their experience with other participants waiting for their turn. Although there is no foolproof plan for such scenarios, we simply asked participants to not return to the waiting area. Similarly, although we limited MTurk workers to participate in our study only once, we do not know whether they were aware of the other conditions since prior work has shown that MTurk workers in India frequently communicate with each other [91]. However, although such strict controls were necessary due to the controlled nature of our experiment, organizations who are simply trying to elicit critical feedback from participants could allow participants to share their experiences, with the expectation that this sharing would increase the amount of social proof that they experience.

Another practical challenge was to determine who to select as confederates. We chose staff members of the grassroots organization because they were available, understood our research, and were trusted by women in the villages. Moreover, since researchers in HCI4D contexts are often accompanied by staff of local organizations who introduce them to communities, we anticipate that other researchers could follow our lead by arranging for local staff to act as confederates. Future work could compare the efficacy of other people playing the role of confederates, such as a health worker or village head.

A key strength of our work is the field evaluation with marginalized women in resource-constrained settings. We designed an intervention to influence response bias and collected strong evidence to demonstrate how social proof could be used to elicit critical feedback on a real, large-scale HCI project deployed in rural India. Although situating our work within the *Projecting Health* project provided several benefits (*e.g.,* access to field locations) there were some disadvantages as well. For example, we were mindful that our intervention must not negatively affect either the *Projecting Health* project or the grassroots organization. For this reason, we rewrote the reviews for the field experiment to ensure that we do not critique key elements of *Projecting Health* such as inviting local people to act in the video and using the local dialect. Other researchers will need to create thoughtful ways to social proof participants without causing damage to existing interventions, local culture and practices, and grassroots organizations.

5.6.3 Ethics

The use of a confederate approach in our experiments introduces important ethical considerations. Specifically, we made participants believe that the reviews they saw had been provided by other participants ‘like them’ when, in reality, we wrote the reviews. This deception was necessary because, for experimental validity, we needed to control the content and length of reviews across conditions. Although we told participants the purpose of the study, we did not tell them about the use of deception. We made this decision after careful thought and prolonged discussion with the *Projecting Health* staff, who thought that disclosure may introduce significant confusion and ultimately cause more harm than good. Although the use of confederates in scientific experiments is well-established in psychology [19, 99], medical [261] and HCI research [106, 121, 155, 133], it should be used with extreme caution. It would be much better to not deceive participants at all. Moving forward, we hope that researchers using our intervention do not need to use deception for the sake of controlling an experiment. Instead, they could seed the intervention with real feedback from real participants and incorporate additional critical feedback into the intervention as it is received.

5.6.4 Limitations

Our work has several limitations. For example, although our intervention is clearly influencing response bias in different ways, it is not necessarily providing researchers with any objective truth. In addition, since we only exposed participants to social proof *before* watching the video, additional research is needed to study the effects of post-exposure. The field experiment was also conducted with a male Indian researcher, and the results may change (*i.e.,* greater bias) if a foreign researcher conducted the evaluation [56]. Moreover, all participants in the field experiment were women, and additional research is needed to examine the impact of ethnicity and gender of researchers and participants on the response bias. Finally, our experiments were done with people in India, in part because prior work [56] demonstrated high-levels of response bias in India. Future research is needed to understand how our results might differ with people in different countries, cultures, and contexts.

CHAPTER 6

SYNTHESIZING AND DRAWING LESSONS FROM CASE STUDIES

Having analyzed the application of my approach (in chapters 3, 4, and 5), I now discuss key takeaways for researchers and practitioners interested in effectively designing technological interventions for behavior change in HCI4D. Each of the case study chapters included a discussion section highlighting contributions to HCI4D and the problem domain studied. In this chapter, I synthesize key themes across all case studies and overall contributions.

A generalizable approach. A major strength of this dissertation is exploring research across diverse problem domains, organization contexts, geographical regions (low-resource and resource-rich countries), and infrastructure. My three case studies took place in Kenya, New York, and India respectively, regions with very different languages, infrastructure, literacy levels, technology access, and cultures. Each case study involved organizations with varying structures, culture, and global reach. Our partners in the first case study involved two NGOs headquartered in East Africa and focused on designing tools for health workers who assist millions of people in hard-to-reach communities. The second case study involved hospitals and three health agencies in New York (including an agency established in 1893). Lastly, the third case study involved an NGO in India that works in 29 countries across five continents. Despite these differences, the same approach was applied with each of these partners to successfully design technological interventions in a way that engaged multiple groups of stakeholders and incorporated Behavioral Science theories. This demonstrates that the approach could be applied when designing technological interventions across a diverse range of HCI4D contexts.

Engaging multiple stakeholder groups in this dissertation and beyond. In engaging multiple stakeholder groups, this dissertation does not intend to replace current HCI4D best practices; rather, it augments existing efforts. The current practice of building partnerships, engaging end users, iterating designs, and evaluating deployments are still core steps that need to be taken during an intervention design that involves a multi-stakeholder approach. Further, researchers could perform multiple steps **in parallel to avoid lengthening their fieldwork research unnecessarily**. Specifically, multiple stakeholders can be identified during field research. Then while preparing for qualitative research to gain a deeper understanding of the contexts of end users, researchers can simultaneously plan to engage other stakeholder groups. This is a more effective approach compared to working sequentially to first engage end users and then later engage other stakeholder groups.

Beyond the three case studies presented, this dissertation approach can also be integrated into other research studies. For example, DeRenzi et al. [57] designed a feedback system for CHWs in India but engaged with only CHWs during their multi-stage design phase. One way to adopt our approach of engaging multiple groups of stakeholders is for the researchers to engage the supervisors of CHWs and the decision makers of the health programs. Early designs can also be shared with these groups, which provides a number of advantages: it reveals important factors that these stakeholder groups are concerned about and could potentially be addressed during early design iterations; it makes the other stakeholder groups feel more connected to the efforts of the research team; and it reduces the possibility of push back when the intervention is about to be scaled.

Similarly, our approach can be applied in the work by Tuli et al. [244] that focused on the design of an SMS system for mothers in Pakistan to improve their knowledge of antenatal care. Specifically, the research team can engage other stakeholder groups such as mothers-in-law and husbands early in the design process. As evidenced in their work, not engaging these groups led to interference in the intervention where the end users, mothers, could not practice the health tips that the SMS application suggested. For example, mothers-in-law sometimes prevented mothers from going to hospitals because they felt it was unnecessary and they gave folklore advice to mothers that conflicted with the advice that these mothers received from doctors. In fact, any HCI4D study that targets end users in an ecosystem where they interact with other stakeholder groups should engage these other groups to create a more effective intervention.

One caveat is if researchers are in an early exploratory phase and have not decided on the specific design space to investigate. For instance, in our third case study, we sought to design a low-cost technique for mitigating participant response bias. After assessing multiple behavioral techniques we converged on *priming* [226] and *social proof* [217]. However, we did not engage any stakeholder groups until we had validated through online experiments that *social proof* is an effective approach for reducing response bias. Only then did we, in our field research in India, engage decision makers interested in eliciting critical feedback from participants in their health program. Engaging stakeholder groups in this way led to meaningful insights for both our researcher team and the health decision makers involved.

It is important to note the subtle difference between engaging decision makers as a stakeholder group and getting "buy-in" from them. In the former, the decision makers are involved in the detail of the intervention while in the latter, they do not engage with the tool—they only know of it and support it. Both steps are necessary, however, without engaging decision makers in the design detail, researchers run a risk of completing the design of a tool that later faces resistance when it is about to be scaled because other stakeholder groups see *"obvious"* design issues that should have been addressed. For instance, in the first case study that involved the design of a health feedback system for care recipients in rural Kenya, our research team shared an early prototype with the director of a county after receiving approval from his office. Upon interacting with the prototype, he requested that the feedback system include an avenue for finding out if health centers delayed treatment to care recipients. Taking his feedback, we carefully designed the system in a way that care recipients could report this specific feedback. After deployment, we summarized preliminary findings and shared them with him, which made him more interested in our project and he recommended that we expand our deployment from one to ten health centers. Without engaging with him during our design, we would not have learned of his concerns or had the opportunity to scale faster. Needless to say, scaling faster would mean more data for research purposes, a highly welcomed benefit, as well as more impact across other communities where care recipients would be enabled to share their own feedback about services received at health centers.

Incorporating behavioral theories in this dissertation and beyond. On the surface, incorporating behavioral theories may appear as merely looking up theories in Behavioral Science literature and selecting a concept that resonates with a researcher. However, this is an illusion as incorporating behavioral theories in all three case studies followed a systematic and thorough process that began at the inception of each study. Each case study started by defining the problem space and the intended behavior to be changed. For example, the first case study focused on *enabling care recipients to engage in a new behavior of sharing feedback*, the second study involved *guiding home health aides in their* *decision making*, and the third study focused on *influencing people's behavior to encourage them to provide critical feedback*. Next, Behavioral Science literature is reviewed to identify potential ways of addressing the problem of interest.

To narrow down from the various behavioral techniques, we focus on behavioral concepts that are **easy to implement and scale through technology**. This step is important because there is little benefit in selecting techniques that are challenging to implement as they will require a lot of resources to scale and this will likely be met with push back from decision makers. Next, the selected techniques are iteratively tested with multiple stakeholder groups until a final version of the intervention is created. In narrowing down to specific behavioral techniques, researchers can select only one technique (e.g. in case study three, we focused on only *social proof* [217] because we wanted a single technique that can be easily adopted by researchers); or multiple techniques as part of a system, where the entire tool is the intervention and the behavioral techniques are supporting components (e.g. the first case study applied *positive reinforcement* [222] to commend users after they submitted feedback and *choice architecture* [11] to simplify the order of the contents that were presented so that users felt the system was easy to use). Following these rigorous steps ensures that behavioral theories are incorporated in a systematic way during the design process of a technological intervention.

The proposed approach for integrating behavioral science theories can also be applied to other HCI4D research beyond the case studies in this dissertation. For example, in the work by DeRenzi et al. [59], the authors used a performance comparison chart that inadvertently discouraged CHWs at the bottom of the chart. However, the researchers could have built on *framing effect* [245] to share this information in a positive way that still challenges CHWs to do more. In their work on using an SMS system to engage men in family planning, Perrier et al. [194] could have tackled low engagement by drawing on *social proof* [217]. To combat the belief where these Kenyan men think they are alone, the SMS messages could have alerted them that *"other men like you are also participating in family planning"*. Further, HCI4D studies that adopt reminder messages to alert users to engage in specific actions [192, 58, 198] could be more intentional about this process by building on *nudge theory* [240] to craft messages that are more persuasive and *choice architecture* [11] to carefully decide when and how user interfaces are designed. It is worth noting that these suggestions do not often require redesigning an intervention from scratch; rather, it involves strengthening existing interventions through carefully crafted content. Therefore, other HCI4D research studies can incorporate Behavioral Science concepts and design more effective technological interventions for under-served communities.

Designing within an existing infrastructure. A key principle adopted in all three case studies was to work within an existing infrastructure in order to ensure project sustainability instead of establishing a system from scratch. As a result, each intervention was designed as a product that could be adopted by our collaborating partners. For instance, in the first case study, the feedback system designed was created in a way that it could be plugged into the mobile health applications used by CHWs. In the second and third case studies, our research team worked with our local partners from the beginning of the project to understand ways in which they could adopt our solution. This approach directly responds to a recent call in HCI4D for researchers to follow an asset based approach that encourages the use of existing infrastructure [191]. It also demonstrates a concrete way for researchers to balance research and practice, a well-known tension in HCI4D [3, 54, 101]. Further, this provides a roadmap for addressing the challenge of sustainability prevalent in HCI4D [256, 192, 250, 260]. For example, by discussing sustainability with our local partners from the beginning of the project, they were interested in how the project could align with their longterm goal of using feedback to improve health services in rural communities. To further make this goal achievable, I took great care in mapping the technical infrastructure of their current systems and shared with them how the new feedback system could be integrated into their existing ecosystem of community health apps. Following these steps encouraged them to take up the cost of expanding the system across several communities after our evaluation showed its potential impact in gathering care recipient feedback. Therefore, it is imperative to work within an existing infrastructure because it promotes sustainability and scalability.

Sustaining rapport with partners. An important lesson for future HCI4D scholars is that a substantial amount of non-research effort was done in parallel with the research activities in order to foster a smooth relationship with our collaborating partners. Sometimes, these efforts were completely unrelated to the research goals, however, they built a stronger connection with partners. For example, during our field research in the first case study, I participated in informal lunch meetings (brown bags) with employees at Medic Mobile [169] where I listened and actively contributed to discussion even when they did not pertain to my field research. During my first encounter with the employees, I was perceived as an *"external researcher from the United States who wouldn't connect with the local employees"* and would be gone after a few weeks of fieldwork. In addition, there were concerns that researchers were arrogant people who looked down on the work of staff. To dispel this myth, I actively inquired about people's work and explained how results of my field research could make their work easier. In scenarios where employees worked on projects unrelated to my field research, I shared insights from related literature on how they could potentially make their work easier. Of course, everyone wants their lives simplified, so this contribution was accepted with open arms. At the end of my stay, many employees in the organization had not only become excited about the potential impact of my field research, but they requested that I return back for another round of fieldwork; and after returning back to the U.S., the stories of increased enthusiasm continued all the way to the leadership of the NGOs who in turn reached out to commend me. Creating this sort of rapport does not happen by sheer luck, rather it takes consistent, conscious efforts. It is tempting to dismiss the need to build rapport as a distraction from core research activities but failing to do could be the end of a research project before it kicks off.

Challenges of adopting a multi-stakeholder view. A major challenge of adopting a multi-stakeholder view is that it implies a substantial increase in the workload of HCI4D researchers. However, this is not the case as this dissertation approach is not a dramatic departure from existing HCI4D practices but a way for HCI4D researchers to go deeper in their work. The HCI4D community already recognizes the essence of tailoring interventions to the needs of end users and conducting iterative designs based on user feedback. In parallel, a growing number of studies have recognized the benefits of identifying all stakeholder groups involved in an ecosystem [18, 138, 244, 181, 184]. This doctoral research builds on this prior work and advocates for the approach proposed, recognizing that failing to do so has real-life consequences that could exacerbate the conditions of marginalized communities. In fact, any researcher looking to conduct quick "parachute" research should avoid HCI4D because it'll likely result in more harm than good. Therefore the question for HCI4D researchers shouldn't be *"is this* *more work?"* but rather *"how can I effectively design my technological intervention?"*

Another challenge is that balancing the views of multiple groups of stakeholders can quickly become contentious during the design process. For example, in our first study, care recipients were, on one hand, concerned that the feedback tool would be difficult for them to use. To address this concern, we simplified the interface and limited the number of survey questions to five. On the other hand, county leaders, as decision makers, wanted to increase the number of questions that recipients answered so they could make data driven decisions on how health programs could improve but this would make the tool difficult to use. Further, CHWs were curious about revealing the identity of people who submitted feedback while care recipients wanted to maintain anonymity during feedback submission. These conflicting issues led to several back and forth communications that on the surface appeared as though we were not making any progress as the same request came up multiple times in subsequent meetings. To tackle this issue, the research team consistently emphasized that adding questions or exposing users' identities would likely discourage care recipients and drive down engagement. In addition, we agreed with decision makers to later extend the questions based on how well users engaged with only five questions in the initial deployment. A key insight in our approach of balancing multiple viewpoints is that it is important **to spend more time in the field** in order to build stronger relationships with stakeholder groups. Further, when balancing views, researchers should **re-align stakeholder groups with the vision of the research project** and highlight the benefits that the communities will experience when the problem is resolved.

Challenges of incorporating behavioral theories. HCI4D scholars interested in incorporating behavioral theories during design may find it challenging to narrow down from the large pool of Behavioral Science concepts to techniques relevant to their research contexts. As a result, this may appear as though one needs to be an expert in Behavioral Science but this is not the case as this dissertation already lays the ground work for future researchers. While it is important to learn Behavioral Science at a deeper level—I learned over the years through multi-day workshops, behavioral science training programs, advanced Ph.D. courses, co-teaching within two years a course on Behavioral Economics for technology, and collaborating with behavioral scientists—one could consult a behavioral science expert during the early exploration phase of a project. For example, in our third case study, we discussed with two behavioral science researchers for insights on potential ways to elicit critical feedback from people. This informed the kinds of search we performed when reviewing behavioral techniques in the literature.

Although there exists several resources for learning about Behavioral Science concepts, I highlight a few (see Appendix A.1 and A.2) that will be relevant to researchers designing for HCI4D contexts: a cognitive bias cheat sheet on human bias [20], a brief introduction to Behavioral Economics [176], and thought-provoking research on the behavior of impoverished populations [50, 23, 160]. While this is not intended to become an exhaustive list, it serves as a primer on integrating behavioral science principles into the design of technological interventions in HCI4D. Therefore, the insights from this doctoral research and the resources shared provide a roadmap for HCI4D researchers to immediately begin to incorporate behavioral science concepts in their technology intervention design.

To assess potential behavioral techniques for technological interventions, it is important to consider if a selected technique is: (i) *practical to implement* using technology; and (ii) *affordable* when scaled across communities for a period of time. For example, designing a tool that builds on social proof [217] requires collecting information on the behavior of end users as well as their peers. However, this could be challenging for a new research project that is yet to collect any data. If the project is eventually set up, it also has to be financially *affordable* during its early phase when the system is still collecting peer data. In such a scenario, researchers need to consider other potential behavioral techniques.

Another key challenge is that **integrating a behavioral lens in HCI4D research does not guarantee behavior change** in end users. Changing human behaviors is difficult because humans are complex beings, and there are social, environmental, and psychological factors at play. Therefore, no one behavioral technique will always effect behavior change. For example, *nudging* patients through SMS reminders to improve medication adherence has been shown to be unsuccessful because adherence is a complex problem that involves deep psychological, sociological, and logistical challenges [170, 218]. As a result, any adopted behavioral technique needs to first be tested before it is scaled across communities. This implies that after going through the process of carefully narrowing down to a behavioral technique, it may not yield the desired effect. Nevertheless, if an intervention does not work after careful design, it still denotes an important outcome for future researchers to build on — a win for Science.

Challenges of success and failure signals. HCI4D researchers have documented the challenges involved in deploying interventions that fail to achieve their original goals [3, 66, 35, 97]. While the definition of *"success"* is still an open debate [54], there are several warning signs that researchers could point to in hindsight that contributed to the failure of a project. In this work, I highlight three signals that have emerged from my work over the years (from the three case studies discussed in this dissertation and from my experience of conducting other research). In Appendix (see A.0.1), I include more questions for researchers to reflect on when conducting fieldwork.

The first signal focuses on experiencing consistent delays from a stakeholder about a crucial task. On the surface, this may seem acceptable—after all, experiencing one form of delay or another is inevitable in HCI4D research. In addition, this behavior doesn't necessarily happen because the stakeholder group concerned is uninterested in the project. On the contrary, they could show a lot of enthusiasm yet not follow through in their actions. For example, in one of my other research projects, the micro-franchisors of a kiosk intervention for young entrepreneurs in Senegal were supposed to check-in weekly on how their entrepreneurs (franchisees) were performing. Our multi-stakeholder engagement with all stakeholder groups revealed that franchisors understood the value of check-ins, had long-term plans of how to support franchisees to grow their small businesses, and assured us that they would do the check-ins "later" in upcoming weeks. Over four weeks later with multiple weekly reminders from our team, the check-ins were still not done. This type of challenge led to a difficult situation where a researcher struggles to call out procrastinative tendencies because they seek to maintain rapport with local partners. However, every excused act translates to time and resources forgone. Attempts to remedy the late weekly check-ins were futile and eventually the project ended after two years of working with these enthusiastic partners. While this was not the only reason for terminating the project, it played a big role. Therefore, when consistent delays are noticed, they should immediately be nipped in the bud by alerting key decision makers to re-strategize as the project is undergoing a pattern that will likely lead to failure.

The second signal centers on the availability of potential participants for the research study. Before going for fieldwork, it is important that a researcher is assured of as many participants as possible that they require for their studies. For example, if one hopes to conduct qualitative research with 20 people, the researcher should ensure that the local partner guarantees that they can recruit more than 20 participants because it is unlikely that all participants will show up. If the local partner cannot guarantee this number of participant pool then the researcher should consider rescheduling their fieldwork, revising their research question or entirely seeking another partner. Having blind faith in enthusiastic local partners is tempting but dangerous because one could spend several months of time and resources in the field, encounter a small number of participants, and ultimately struggle to do meaningful work that could be shared or published in their research communities. Needless to say, time spent performing such field work has grave consequences such as delaying one's research progress or graduation from a doctoral program.

The final signal primarily focuses on the researcher leading the project and it involves having at least one back up plan before arriving in the field. Fieldwork rarely goes as one anticipates and going into the field without detailed alternative plans is a recipe for disaster. For example, in our first study on designing a feedback tool for care recipients, we considered deploying an IVR system but learned during fieldwork that approval to use IVR through a telecommunication network involves a bureaucratic process that would take several weeks. As a result, we defaulted to our alternative plan to use a USSD based tool as a backup option. In addition, we planned to test a prototype system on participants mobile devices but anticipated that they may not have their own personal phones so we bought a spare mobile device that we could use for testing. However, we ended up not needing this as most participants had their own devices. We also learned during initial focus group discussions that participants interacted a lot through SMS. Building on this insight, we immediately developed an SMS reminder as part of our intervention tool because the network API had been studied beforehand during our planning stage. Without making detailed plans ahead of time, one risks performing a lot of work yet feeling lost in the unpredictable flow of fieldwork. In sharing these three indicators, my hope is that they will save researchers the time and resources much needed in their work, as well as provide a sense of early warning signals about a project going in the wrong direction.

Future directions. Adopting the approach described in this dissertation is not intended to become the be-all-end-all for designing HCI4D interventions. No one approach is. However, this dissertation serves as pioneering work that opens up a new space on how to effectively design technological interventions in HCI4D. It also reignites the conversation on the extent to which research should meddle with practice: *should research focus more on development 'outputs' (short-term goals) and less on 'outcomes' (long-term results) [54]*? Regardless of the stance of HCI4D researchers and practitioners, it is important to remember that these decisions impact the lives of people who are under-served, under-resourced, and under-represented.

The HCI4D community has made and continues to make substantial strides in catering to under-served communities in the world. Although this is well-known, it is worth reiterating that there are other research communities that share similar agendas but have seen little collaboration. For example, the field of Behavioral Economics for Development has made notable strides in advancing work in under-served communities including in the areas of health [13, 224], finance [10, 31], and education [12, 78]. Notable researchers (such as Sendhil

Mullainathan [175], Eldar Shafir [216], and Johannes Haushofer [96]) have also written about advancing development efforts. Herein lies an opportunity for HCI4D researchers to collaborate.

Conversely, behavioral researchers outside HCI4D interested in integrating technology into their work can collaborate with HCI4D researchers to create effective technological interventions that adopt a bottom-up, participatory design approach instead of the current method of a top-down, one-size-fits-all approach that ignores local community contexts. To foster collaborations, Behavioral Science scholars could attend premier HCI4D conferences such as CHI, CSCW, ICTD; and HCI4D researchers could in turn attend Behavioral Science conferences such as BSPA, SJDM, and APS to connect with other researchers and practitioners. Bridging both disciplines would give HCI4D much needed visibility and expand the horizon of the larger HCI discipline. In parallel, it would answer the call for behavioral economists focusing on international development to integrate their work with that of technologists [50]. In so doing, the bridge between HCI4D and Behavioral Science will become further strengthened with more opportunities to design interventions that make the world a better place for under-served communities.

CHAPTER 7

CONCLUSION

This dissertation focuses on how to advance current best practices for designing technologies in HCI4D by: (i) adopting a multi-stakeholder approach that engages other stakeholder groups beyond the end users; and (ii) systematically incorporating behavioral science theories into the design process to reduce the possibility of encountering unintended consequences. Through three case studies, I lay the foundation for how researchers can better design technologies for under-served communities. In the first case study, I focused on the design of a feedback system for care recipients to share feedback about the health services they receive in rural Kenya. In the second case study, I explored the design of a technological intervention for home health aides supporting heart failure patients in New York City. In the third case study, I presented the design of a low-cost, behavioral intervention for researchers to mitigate participant response bias in rural India. By adopting the same approach across different problem domains, organization contexts, geographical regions (low-resource and resource-rich countries), and infrastructure, I demonstrate that this dissertation approach could be applied across a diverse range of HCI4D contexts. In doing so, this doctoral research makes multiple contributions to the HCI4D community through: the findings of each of the three case studies, providing a roadmap for future scholars to design digital interventions more effectively in HCI4D, and builds a stronger bridge between the fields of HCI4D and Behavioral Science.

APPENDIX A

REFLECTION FOR FIELDWORK AND BEHAVIORAL THEORIES

Conducting fieldwork is a challenging process. Even when done properly, there is no guarantee that a project will become successful. Over the years, I have generated a few questions and signals that have helped in successfully designing interventions for under-served communities. Below, I summarize these questions for HCI4D researchers and practitioners, particularly scholars new to HCI4D, to reflect on during the design of their own technological interventions.

A.0.1 Questions to Reflect on for Fieldwork Research

Part A: key questions when adopting a multi-stakeholder approach

1. Has the research team identified primary stakeholder groups in the ecosystem of interest?
2. During the iterative design process, did you engage with other stakeholder groups beyond the end users of the intervention?
3. Have you made plans to engage stakeholder groups in parallel (during the same fieldwork) instead of engaging with groups sequentially (multiple rounds of fieldwork that will delay the design process unnecessarily)?

Part B: key questions when incorporating behavioral theories

1. Has the research team defined the specific user behavior they intend to change? Example: to eat healthier, to save more money, etc.
2. Have you reviewed research on cognitive biases that under-served communities experience?

3. Have you reviewed the list of behavioral concepts that could potentially apply to your research context?
4. To narrow down to specific behavioral techniques to adopt, have you checked if the technique is (i) practical to implement with technology; and (ii) affordable if it is scaled to a larger group of people?
5. Have you reflected on the potential of the selected technique backfiring in the community or only working in the short term? Although it is challenging to know before hand how an intervention will turn out, in specific contexts, some interventions are more likely than others to lead to harmful outcomes.

Part C: key questions to become more productive during HCI4D fieldwork

1. Has the research team created at least one back up plan before going for fieldwork?
2. Have local partners guaranteed the research team that there will be enough participants for the fieldwork? This is particularly important because it is common for participants to not show up.
3. During field work, does the research team recognize early when stakeholders are consistently delaying on carrying out crucial tasks as this is a signal that the project is not going well? It is necessary to perform this check even if the reasons for delays are valid and local partners are enthusiastic about the project. Delays can lead to late project deployment, no research publications, set back in graduation from doctoral program, etc.
4. Instead of creating the intervention as an adhoc tool, will the intervention be designed as a tool within an existing infrastructure as this approach

facilitates scaling and adoption? Example: design a feedback tool in a way that it can be integrated into existing mobile apps for health workers.

5. Has the research team started conversations about scaling and sustainability during the inception of the project and not after deployment? This step is beneficial even if the researchers are not looking to scale the project because insights from conversations could feature lead to design recommendations.

A.0.2 Behavioral Concepts Relevant to HCI4D Interventions

There are several behavioral science theories to choose from when designing tools for HCI4D contexts. Table A.1 and Table A.2 show behavioral concepts with examples that serve as a good starting point when considering behavioral theories to choose from. Each example focuses on a specific domain such as health, education or finance. However, the behavioral concepts are not restricted to only the example domain and can be applied to any other contexts. Although all examples are not strictly in HCI4D contexts, they provide insights that can be potentially adopted in HCI4D.

Behavioral Concept	**Description**
A brief introduction to Behavioral Economics [210, 176]	Behavioral Economics principles are important for gaining a better understanding of how to design effective interventions. Although the referenced articles do not focus on HCI4D contexts, they provide an introduction to fundamental concepts important for incorporating behavioral theories. For the article by Samson and Ariely [210], I recommend reading at least the first 20 pages.
A brief introduction to cognitive biases [20]	Understanding cognitive biases provides insights into why people behave in irrational ways and this knowledge can inform the design of technological interventions. The referenced article is a 10-minute read that you can always come back to.
A brief introduction to the behavior of impoverished populations [50, 23, 160]	These articles provide a deep understanding into designing interventions (mostly non-technical) for users in low- and middle-income countries. Insights from these papers can be adapted to technological contexts in HCI4D.
Boosting and Self-Awareness [43, 98]	This involves providing users with information that increases their level of self awareness. Example: As a small business owner, you saved 10 shillings last week compared to your goal of saving 1000 shillings.
Choice architecture [11, 117]	The order of presentation of items impacts people's decision making. For example, a digital feedback tool shows users simple questions first (e.g. your gender) before asking them to later reflect on the kind of feedback they want to share. Interacting with the easy questions first make users feel that the tool is indeed easy to use.
Framing effect [126, 245]	Presenting the same information in a positive or negative way to impact users' behaviors. Example: "*You have completed only 20% of practice quizzes in preparation for your high school exam next month*" (frames user behavior positively) vs. "*You have not completed 80% of practice quizzes in preparation for your high school exam next month*" (frames user behavior negatively).
Gamification [88, 116]	Assigning non-monetary rewards to boost motivation. Example: You will earn a badge of "health worker leader" after you complete your next household visit.
Goal Setting [59, 235, 153]	Designing an action plan that motivates one to achieve a personal target. Example: a health worker sets a personal goal to visit the households of 10 clients in one week.

Table A.1: Behavioral concepts to consider when designing technological interventions for under-served communities. Each example can be applied beyond the example context and potentially tailored to HCI4D settings.

Behavioral Concept	**Description**
Incentives [85, 6, 123]	Monetary benefits to increase motivation. Example: you will qualify for a raffle draw of 10 minutes of airtime if you attend all your classes this month.
Intention-Action Gap [249, 87]	A behavioral bias observed when one intends to engage in an activity but do not follow through in action. Example: Planning at the beginning of a new year to go to the gym frequently (intentions) but never visiting the gym after several months (action).
Loss Aversion [90, 122]	People value losses at two times its magnitude than they value gains. Example: As a student, you will lose 15 shillings of airtime today if you do not attend classroom lectures.
Negative Reinforcement [182, 112, 221]	Strengthening the likelihood of a behavior happening again in the future by making users avoid a negative outcome or aversive stimulus. Example: You finish your homework on time (behavior) to avoid a repetitive phone reminder that your homework hasn't been completed (aversive stimulus).
Nudge [183, 240]	Steering users in a specific direction that is beneficial to the user without forcing them to make a specific choice. Example: If you save only 10 shillings today, you will be able to purchase textbooks for your child at the end of the month.
Pre-commitment [7, 214]	Leverage cognitive bias that humans tend to think optimistically of their future selves. Example question at the beginning of a month: As a small business owner, commit now to the amount you plan to save this month?
Present Bias [179, 71]	A cognitive bias that makes one take actions for immediate pleasure instead of waiting for a larger future reward. Example: Spend your monthly salary buying new clothes instead of saving it towards retirement.
Positive Reinforcement [227, 221]	Rewarding a behavior with a reinforcing stimulus in order to increase the likelihood of future repetitions. Example: You receive 50 rupees of airtime (reinforcing stimulus) for every passing grade you achieve at the end of a semester (behavior).
Social Proof [255, 61, 217]	Influence a user's action by revealing the behaviors of others who share similar characteristics. Example: 50 entrepreneurs from your village saved over 100 rupees this week. You should join them and save your money too.

Table A.2: More behavioral concepts to consider when designing technological interventions for under-served communities. Each example can be applied beyond the example context and potentially tailored to HCI4D settings.

BIBLIOGRAPHY

[1] Amazon. Mechanical Turk. https://www.mturk.com, 2017. [Online; accessed September 10, 2017].

[2] Andrew P Ambrosy, Gregg C Fonarow, Javed Butler, Ovidiu Chioncel, Stephen J Greene, Muthiah Vaduganathan, Savina Nodari, Carolyn SP Lam, Naoki Sato, Ami N Shah, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *Journal of the American College of Cardiology*, 63(12):1123–1133, 2014.

[3] Yaw Anokwa, Thomas N Smyth, Divya Ramachandran, Jahanzeb Sherwani, Yael Schwartzman, Rowena Luk, Melissa Ho, Neema Moraveji, and Brian DeRenzi. Stories from the field: Reflections on hci4d experiences. *Information Technologies & International Development*, 5(4):pp–101, 2009.

[4] Anthony Dunne and Fionna Raby. *Speculative Everything*. 2013.

[5] Ajit Appari, M Eric Johnson, and Denise L Anthony. Hipaa compliance in home health: a neo-institutional theoretic perspective. In *Proceedings of the first ACM workshop on Security and privacy in medical and home-care systems*, pages 13–20. ACM, 2009.

[6] Dan Ariely, Anat Bracha, and Stephan Meier. Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–55, 2009.

[7] Dan Ariely and Klaus Wertenbroch. Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science*, 13(3):219–224, 2002.

[8] Alison Ashley, Sandra S Butler, and Nancy Fishwick. Home-care-aides voices from the field: Job experiences of personal support specialists — the maine home care worker retention study. *Home healthcare nurse*, 28(7):399, 2010.

[9] Nava Ashraf, Oriana Bandiera, and B Kelsey Jack. No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17, 2014.

[10] Nava Ashraf, Dean Karlan, and Wesley Yin. Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121(2):635–672, 2006.

[11] J Balz, C Sunstein, and R Thaler. Choice architecture. *E. Shafir, The behavioral foundations of public policy*, pages 428–439, 2014.

[12] Abhijit V Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden. Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

[13] Abhijit Vinayak Banerjee, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. Improving immunisation coverage in rural india: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *Bmj*, 340:c2220, 2010.

[14] Abdullah H Baqui, Shams El-Arifeen, Gary L Darmstadt, Saifuddin Ahmed, Emma K Williams, Habibur R Seraji, Ishtiaq Mannan, Syed M Rahman, Rasheduzzaman Shah, Samir K Saha, et al. Effect of community-based newborn-care intervention package implemented through two service-delivery strategies in sylhet district, bangladesh: a cluster-randomised controlled trial. *The lancet*, 371(9628):1936–1944, 2008.

[15] Shaowen Bardzell. Feminist HCI: Taking Stock and Outlining an Agenda for Design. Technical report, 2010.

[16] Shaowen Bardzell and Jeffrey Bardzell. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 675–684. ACM, 2011.

[17] Peter Barron, Yogan Pillay, Antonio Fernandes, Jane Sebidi, and Rob Allen. The momconnect mhealth initiative in south africa: Early impact on the supply side of mch services. *Journal of public health policy*, 37(2):201–212, 2016.

[18] Amna Batool, Samia Razaq, Maham Javaid, Beenish Fatima, and Kentaro Toyama. Maternal complications: Nuances in mobile interventions for maternal health in urban pakistan. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, page 3. ACM, 2017.

[19] Diana Baumrind. Research using intentional deception: ethical issues revisited. *American psychologist*, 40(2):165, 1985.

[20] Buster Benson. Cognitive bias cheat sheet, 2016. Retrieved October 29, 2019 from `https://medium.com/better-humans/cognitive-bias-cheat-sheet-55a472476b18`.

[21] Anita Bercovitz, Abigail J Moss, Manisha Sengupta, Eunice Y Park-Lee, Adrienne Jones, Lauren D Harris-Kojetin, and Marie R Squillace. *An overview of home health aides: United States, 2007*. Citeseer, 2008.

[22] Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 4chan and/b: An analysis of

anonymity and ephemerality in a large online community. In *ICWSM*, pages 50–57, 2011.

[23] Marianne Bertrand, Sendhil Mullainathan, and Eldar Shafir. A behavioral-economics view of poverty. *American Economic Review*, 94(2):419–423, 2004.

[24] Nicola J Bidwell, Thomas Reitmaier, Gary Marsden, and Susan Hansen. Designing with mobile digital storytelling in rural africa. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1593–1602. ACM, 2010.

[25] Julian Bleecker. Design Fiction: A Short Essay on Design, Science, Fact and Fiction. *Near Future Laboratory*, (March):49, 2009.

[26] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[27] Rod Bond and Peter B Smith. Culture and conformity: A meta-analysis of studies using asch's (1952b, 1956) line judgment task. *Psychological bulletin*, 119(1):111, 1996.

[28] Eric Brewer, Michael Demmer, Bowei Du, Melissa Ho, Matthew Kam, Sergiu Nedevschi, Joyojeet Pal, Rabin Patra, Sonesh Surana, and Kevin Fall. The case for technology in developing regions. *Computer*, 38(6):25–38, 2005.

[29] Barry Brown, Stuart Reeves, and Scott Sherwood. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1657–1666. ACM, 2011.

[30] Suzana Brown and Alan Mickelson. Why some well-planned and community-based ictd interventions fail. *Information Technologies & International Development*, 15:13, 2019.

[31] Lasse Brune, Xavier Giné, Jessica Goldberg, and Dean Yang. *Commitments to save: A field experiment in rural Malawi*. The World Bank, 2011.

[32] Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, and Gaetano Borriello. Open data kit 2.0: expanding and refining information services for developing regions. In *Proceedings of the 14th workshop on mobile computing systems and applications*, page 10. ACM, 2013.

[33] Bureau of Labor Statistics USDoL. Home Health Aides and Personal Care Aides, 2016. Retrieved Dec 24, 2018 from `https://www.bls.gov/ooh/healthcare/home-health-aides-and-personal-care-aides.htm`.

[34] Moira Burke, Cameron Marlow, and Thomas Lento. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 945–954. ACM, 2009.

[35] Jenna Burrell and Kentaro Toyama. What constitutes good ICTD research? *Information Technologies & International Development*, 5(3):pp–82, 2009.

[36] Dipanjan Chakraborty and Aaditeshwar Seth. Building citizen engagement into the implementation of welfare schemes in rural india. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, page 22. ACM, 2015.

[37] Apala Lahiri Chavan. Another culture, another method. In *Proceedings of the 11th International Conference on Human-Computer Interaction*, volume 21. Citeseer, 2005.

[38] Richard K Cheng, Margueritte Cox, Megan L Neely, Paul A Heidenreich, Deepak L Bhatt, Zubin J Eapen, Adrian F Hernandez, Javed Butler, Clyde W Yancy, and Gregg C Fonarow. Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the medicare population. *American heart journal*, 168(5):721–730, 2014.

[39] Marshini Chetty and Rebecca E Grinter. Hci4d: hci challenges in the global south. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2327–2332. ACM, 2007.

[40] Robert B Cialdini, Wilhelmina Wosinska, Daniel W Barrett, Jonathan Butner, and Malgorzata Gornik-Durose. Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, 25(10):1242–1253, 1999.

[41] Pamela F Cipriano, Kathryn Bowles, Maureen Dailey, Patricia Dykes, Gerri Lamb, and Mary Naylor. The importance of health information technology in care coordination and transitional care. *Nursing outlook*, 61(6):475–489, 2013.

[42] Community Health Toolkit, 2019. Retrieved April 4, 2019 from `https://communityhealthtoolkit.org/`.

[43] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1797–1806. ACM, 2008.

[44] Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. Is seeing believing?: how recommender system interfaces affect

users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 585–592. ACM, 2003.

[45] Steven S Coughlin. Recall bias in epidemiologic studies. *Journal of clinical epidemiology*, 43(1):87–91, 1990.

[46] Jean Czerlinski, Gerd Gigerenzer, and Daniel G Goldstein. How good are simple heuristics? In *Simple heuristics that make us smart*, pages 97–118. Oxford University Press, 1999.

[47] Abdi Latif Dahir. Airtel and telkom are uniting to take on safaricom in kenya, 2019. Retrieved March 26, 2019 from `http://bit.ly/takeOnSafaricom`.

[48] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 739–749. ACM, 2014.

[49] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. The role of social influence in security feature adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1416–1426. ACM, 2015.

[50] Saugato Datta and Sendhil Mullainathan. Behavioral design: a new approach to development policy. *Review of Income and Wealth*, 60(1):7–35, 2014.

[51] Sophie Sahaf Reem Rahman David Bonbright, Britt Lake and Renee Ho. Net promoter score for the nonprofit sector: What we've learned so far, 2015. Retrieved March 18, 2019 from `http://bit.ly/npsScore`.

[52] Steven L Dawson and Rick Surpin. The home health aide: scarce resource in a competitive marketplace. *Care Management Journals*, 2(4):226–231, 2000.

[53] Edward L Deci, Richard Koestner, and Richard M Ryan. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627, 1999.

[54] Nicola Dell and Neha Kumar. The ins and outs of hci for development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2220–2232. ACM, 2016.

[55] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. ”yours is better!”: Participant response bias in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’12, pages 1321–1330, New York, NY, USA, 2012. ACM.

[56] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. Yours is better!: participant response bias in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1321–1330. ACM, 2012.

[57] Brian DeRenzi, Nicola Dell, Jeremy Wacksman, Scott Lee, and Neal Lesh. Supporting community health workers in india through voice-and web-based feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2770–2781. ACM, 2017.

[58] Brian DeRenzi, Leah Findlater, Jonathan Payne, Benjamin Birnbaum, Joachim Mangilima, Tapan Parikh, Gaetano Borriello, and Neal Lesh. Improving community health worker performance through automated sms. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*, pages 25–34. ACM, 2012.

[59] Brian DeRenzi, Jeremy Wacksman, Nicola Dell, Scott Lee, Neal Lesh, Gaetano Borriello, and Andrew Ellner. Closing the feedback loop: A 12-month evaluation of asta, a self-tracking application for ashas. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, page 22. ACM, 2016.

[60] Pooja M. Desai, Matthew E. Levine, David J. Albers, and Lena Mamykina. Pictures worth a thousand words: Reflections on visualizing personal blood glucose forecasts for individuals with type 2 diabetes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 538:1–538:13, New York, NY, USA, 2018. ACM.

[61] Morton Deutsch and Harold B Gerard. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629, 1955.

[62] DHIS2. District Health Information Software 2, 2019. Retrieved April 4, 2019 from `https://www.dhis2.org/`.

[63] Francis J Di Vesta. Effects of confidence and motivation on susceptibility to informational social influence. *The Journal of Abnormal and Social Psychology*, 59(2):204, 1959.

[64] Tawanna R Dillahunt. Fostering social capital in economically distressed communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 531–540. ACM, 2014.

[65] Dimagi. Commcare by dimagi, 2019. Retrieved March 11, 2019 from `https://www.dimagi.com/commcare/`.

[66] Leslie L Dodson, S Sterling, and John K Bennett. Considering failure: eight years of itid research. In *Proceedings of the Fifth International Conference on*

Information and Communication Technologies and Development, pages 56–64. ACM, 2012.

[67] Lynn Dombrowski, Adriana Alvarado Garcia, and Jessica Despard. Low-wage precarious workers' sociotechnical practices working towards addressing wage theft. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2017-May, pages 4585–4598, 2017.

[68] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. Social justice-oriented interaction design: Outlining key design strategies and commitments. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 656–671. ACM, 2016.

[69] Andy Dow, John Vines, Rob Comber, and Rob Wilson. Thoughtcloud: Exploring the role of feedback technologies in care organisations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3625–3636. ACM, 2016.

[70] Andy Dow, John Vines, Toby Lowe, Rob Comber, and Rob Wilson. What happens to digital feedback?: Studying the use of a feedback capture platform by care organisations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5813–5825. ACM, 2017.

[71] Esther Duflo, Michael Kremer, and Jonathan Robinson. Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American economic review*, 101(6):2350–90, 2011.

[72] Olive Jean Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.

[73] Olive Jean Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.

[74] Jordan Eschler, Logan Kendall, Kathleen O'Leary, Lisa M. Vizer, Paula Lozano, Jennifer B. McClure, Wanda Pratt, and James D. Ralston. Shared calendars for home health management. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '15, pages 1277–1288, New York, NY, USA, 2015. ACM.

[75] Jordan Eschler and Wanda Pratt. "i'm so glad i met you": Designing dynamic collaborative support for young adult cancer survivors. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1763–1774, New York, NY, USA, 2017. ACM.

[76] Brittany Fiore-Silfvast, Carl Hartung, Kirti Iyengar, Sharad Iyengar, Kiersten Israel-Ballard, Noah Perin, and Richard Anderson. Mobile video for patient education: the midwives' perspective. In *Proceedings of the 3rd ACM Symposium on Computing for Development*, page 2. ACM, 2013.

[77] Emily Franzosa, Emma K Tsui, and Sherry Baron. Home health aides' perceptions of quality care: Goals, challenges, and implications for a rapidly changing industry. *NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy*, 27(4):629–647, 2018.

[78] Roland G Fryer Jr. Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4):1755–1798, 2011.

[79] Adrian Furnham. Response bias, social desirability and dissimulation. *Personality and individual differences*, 7(3):385–400, 1986.

[80] Mirta Galesic and Michael Bosnjak. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, 73(2):349–360, 2009.

[81] Mrunal Gawade, Rajan Vaish, Mercy Nduta Waihumbu, and James Davis. Exploring employment opportunities through microtasks via cybercafes. In *Global Humanitarian Technology Conference (GHTC), 2012 IEEE*, pages 77–82. IEEE, 2012.

[82] Jeffrey Gettleman. Meant to Keep Malaria Out, Mosquito Nets Are Used to Haul Fish In, 2015. Retrieved October 27, 2019 from `http://bit.ly/nyt-mosquito`.

[83] Matlin Gilman, Jason M Hockenberry, E Kathleen Adams, Arnold S Milstein, Ira B Wilson, and Edmund R Becker. The financial effect of value-based purchasing and the hospital readmissions reduction program on safety-net hospitals in 2014. *Annals of Internal Medicine*, 163(6):427–436, 2015.

[84] Mareike Glöss, Moira McGregor, and Barry Brown. Designing for labour: uber and the on-demand mobile workforce. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1632–1643. ACM, 2016.

[85] Uri Gneezy and Aldo Rustichini. Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810, 2000.

[86] John Gerard Scott Goldie. Connectivism: A knowledge learning theory for the digital age? *Medical teacher*, 38(10):1064–1069, 2016.

[87] Peter M Gollwitzer. Implementation intentions: strong effects of simple plans. *American psychologist*, 54(7):493, 1999.

[88] Daniel Gooch, Asimina Vasalou, Laura Benton, and Rilla Khaled. Using gamification to motivate students with dyslexia. In *Proceedings of the 2016 CHI Conference on human factors in computing systems*, pages 969–980. ACM, 2016.

[89] Shion Guha, Mohammad Rashidujjaman Rifat, Faysal Hossain Shezan, Nicola Dell, et al. Privacy vulnerabilities in the practices of repairing broken digital artifacts in bangladesh. *Information Technologies and International Development*, 2017.

[90] Junius Gunaratne and Oded Nov. Informing and improving retirement saving performance using behavioral economics theory-driven user interfaces. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 917–920. ACM, 2015.

[91] Neha Gupta, David Martin, Benjamin V. Hanrahan, and Jacki O'Neill. Turk-life in india. In *Proceedings of the 18th International Conference on Supporting Group Work*, GROUP '14, pages 1–11, 2014.

[92] DF Hamilton, Judith V Lane, P Gaston, JT Patton, DJ Macdonald, AHRW Simpson, and CR Howie. Assessing treatment outcomes using a single question: the net promoter score. *The bone & joint journal*, 96(5):622–628, 2014.

[93] Michael D Hanus and Jesse Fox. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & education*, 80:152–161, 2015.

[94] Steve Haroz, Robert Kosara, and Steven L Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 1191–1200. ACM, 2015.

[95] Lane Harrison, Katharina Reinecke, and Remco Chang. Infographic aesthetics: Designing for the first impression. In *Proceedings of the 33rd Annual*

ACM Conference on Human Factors in Computing Systems, pages 1187–1190. ACM, 2015.

[96] Johannes Haushofer. Johannes haushofer, 2019. Retrieved November 16, 2019 from `https://psych.princeton.edu/person/eldar-shafir/`.

[97] Richard Heeks. The ict4d 2.0 manifesto: Where next for icts and international development? *Development Informatics working paper*, (42), 2009.

[98] Ralph Hertwig and Till Grüne-Yanoff. Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6):973–986, 2017.

[99] Ralph Hertwig and Andreas Ortmann. Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1):59–92, 2008.

[100] Sarah J Hewko, Sarah L Cooper, Hanhmi Huynh, Trish L Spiwek, Heather L Carleton, Shawna Reid, and Greta G Cummings. Invisible no more: a scoping review of the health care aide workforce literature. *BMC nursing*, 14(1):38, 2015.

[101] Melissa R Ho, Thomas N Smyth, Matthew Kam, and Andy Dearden. Human-computer interaction for development: The past, present, and future. *Information Technologies & International Development*, 5(4):pp–1, 2009.

[102] Isaac Holeman, Tara Patricia Cookson, and Claudia Pagliari. Digital technology for health sector governance in low and middle income countries: a scoping review. *Journal of global health*, 6(2), 2016.

[103] Isaac Holeman and Dianna Kane. Human-centered design for global health equity. *Information Technology for Development*, forthcoming.

[104] Nick Hughes and Susie Lonie. M-pesa: mobile money for the unbanked turning cellphones into 24-hour tellers in kenya. *Innovations: technology, governance, globalization*, 2(1-2):63–81, 2007.

[105] Jessica Hullman, Eytan Adar, and Priti Shah. The impact of social information on visual judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1461–1470. ACM, 2011.

[106] Jennifer Hyde, Sara Kiesler, Jessica K Hodgins, and Elizabeth J Carter. Conversing with children: cartoon and video people elicit similar conversational behaviors. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1787–1796. ACM, 2014.

[107] Institute PHC. America's Direct Care Workforce, 2013. Retrieved Dec 24, 2018 from `https://phinational.org/wp-content/uploads/legacy/phi-facts-3.pdf`.

[108] Lilly Irani. Hci on the move: methods, culture, values. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 2939–2942. ACM, 2010.

[109] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. Postcolonial computing. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1311, New York, New York, USA, 2010. ACM Press.

[110] Lilly C Irani and M. Six Silberman. Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 611–620, 2013.

[111] Azra Ismail, Naveena Karusala, and Neha Kumar. Bridging disconnected knowledges for community health. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):75, 2018.

[112] Brian A Iwata. Negative reinforcement in applied behavior analysis: An emerging technology. *Journal of Applied Behavior Analysis*, 20(4):361–378, 1987.

[113] Helen E Jack, Sophia D Arabadjis, Lucy Sun, Erin E Sullivan, and Russell S Phillips. Impact of community health workers on use of healthcare services in the united states: a systematic review. *Journal of general internal medicine*, 32(3):325–344, 2017.

[114] Maia Jacobs, James Clawson, and Elizabeth D. Mynatt. Cancer navigation: Opportunities and challenges for facilitating the breast cancer journey. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14, pages 1467–1478, New York, NY, USA, 2014. ACM.

[115] Stephen F Jencks, Mark V Williams, and Eric A Coleman. Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428, 2009.

[116] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2001–2013. ACM, 2016.

[117] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2):487–504, 2012.

[118] Christine D Jones, Adit A Ginde, Robert E Burke, Heidi L Wald, Frederick A Masoudi, and Rebecca S Boxer. Increasing home healthcare referrals upon

discharge from us hospitals: 2001–2012. *Journal of the American Geriatrics Society*, 63(6):1265–1266, 2015.

[119] Christine D Jones, Heidi L Wald, Rebecca S Boxer, Frederick A Masoudi, Robert E Burke, Roberta Capp, Eric A Coleman, and Adit A Ginde. Characteristics associated with home health care referrals at hospital discharge: results from the 2012 national inpatient sample. *Health services research*, 52(2):879–894, 2017.

[120] HIPAA Journal. Hipaa compliant checklist, 2018. Retrieved Nov 16, 2018 from `http://bit.ly/hipaaChecklist`.

[121] Hee-Tae Jung, Yu-kyong Choe, and Roderic Grupen. Extended virtual presence of therapists through home service robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 209–210. ACM, 2015.

[122] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206, 1991.

[123] Emir Kamenica. Behavioral economics and psychology of incentives. *Annu. Rev. Econ.*, 4(1):427–452, 2012.

[124] Shreya Kangovi, Nandita Mitra, Lindsey Norton, Rory Harte, Xinyi Zhao, Tamala Carter, David Grande, and Judith A Long. Effect of community health worker support on clinical outcomes of low-income patients across primary care facilities: a randomized clinical trial. *JAMA Internal Medicine*, 2018.

[125] Elizabeth Kaziunas, Ayse G. Buyuktur, Jasmine Jones, Sung W. Choi, David A. Hanauer, and Mark S. Ackerman. Transition and reflection

in the use of health information: The case of pediatric bone marrow transplant caregivers. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work; Social Computing*, CSCW ’15, pages 1763–1774, New York, NY, USA, 2015. ACM.

[126] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 272–283. ACM, 2016.

[127] Aniket Kittur, Ed H Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM, 2008.

[128] Rebecca Koladycz, Gwendolyn Fernandez, Kate Gray, and Heidi Marriott. The net promoter score (nps) for insight into client experiences in sexual and reproductive health clinics. *Global Health: Science and Practice*, 6(3):413–424, 2018.

[129] Peter Kollock. The production of trust in online markets. *Advances in group processes*, 16(1):99–123, 1999.

[130] Steven Komarov, Katharina Reinecke, and Krzysztof Z Gajos. Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 207–216. ACM, 2013.

[131] Stephen Kosack and Archon Fung. Does transparency improve governance? *Annual Review of Political Science*, 17:65–87, 2014.

[132] David Kotz, Sasikanth Avancha, and Amit Baxi. A privacy framework for mobile health and home-care systems. In *Proceedings of the first ACM*

workshop on Security and privacy in medical and home-care systems, pages 1–12. ACM, 2009.

[133] Peter M. Krafft, Michael Macy, and Alex ”Sandy” Pentland. Bots as virtual confederates: Design and ethics. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’17, pages 183–190, 2017.

[134] Adam DI Kramer. The spread of emotion via facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 767–770. ACM, 2012.

[135] Harlan M Krumholz, Angela R Merrill, Eric M Schone, Geoffrey C Schreiner, Jersey Chen, Elizabeth H Bradley, Yun Wang, Yongfei Wang, Zhenqiu Lin, Barry M Straube, et al. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. *Circulation: Cardiovascular Quality and Outcomes*, 2(5):407–413, 2009.

[136] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[137] Neha Kumar and Richard J Anderson. Mobile phones for maternal health in rural india. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 427–436. ACM, 2015.

[138] Neha Kumar, Waylon Brunette, Nicola Dell, Trevor Perrier, Beth Kolko, Gaetano Borriello, and Richard Anderson. Understanding sociotechnical implications of mobile health deployments in india, kenya, and zimbabwe. *Information Technologies & International Development*, 11(4), 2015.

[139] Neha Kumar, Nassim Jafarinaimi, and Mehrab Bin Morshed. Uber in Bangladesh. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–21, 11 2018.

[140] Neha Kumar, Trevor Perrier, Michelle Desmond, Kiersten Israel-Ballard, Vikrant Kumar, Sudip Mahapatra, Anil Mishra, Shreya Agarwal, Rikin Gandhi, Pallavi Lal, et al. Projecting health: Community-led video education for maternal health. In *The International Conference on Information and Communication Technologies and Development*, page 17. ACM, 2015.

[141] Vishwajeet Kumar, Saroj Mohanty, Aarti Kumar, Rajendra P Misra, Mathuram Santosham, Shally Awasthi, Abdullah H Baqui, Pramod Singh, Vivek Singh, Ramesh C Ahuja, et al. Effect of community-based behaviour change management on neonatal mortality in shivgarh, uttar pradesh, india: a cluster-randomised controlled trial. *The Lancet*, 372(9644):1151–1162, 2008.

[142] Richard E. Ladner. Design for user empowerment. *Interactions*, 22(2):24–29, 3 2015.

[143] Mars Lan, Lauren Samy, Nabil Alshurafa, Myung-Kyung Suh, Hassan Ghasemzadeh, Aurelia Macabasco-O'Connell, and Majid Sarrafzadeh. Wanda: An end-to-end remote health monitoring and analytics system for heart failure patients. In *Proceedings of the Conference on Wireless Health*, WH '12, pages 9:1–9:8, New York, NY, USA, 2012. ACM.

[144] Marc Langheinrich. Privacy by design — principles of privacy-aware ubiquitous systems. In *International conference on Ubiquitous Computing*, pages 273–291. Springer, 2001.

[145] Shaimaa Lazem and Hussein Aly Jad. We play we learn: Exploring the value of digital educational games in rural egypt. In *Proceedings of the 2017*

CHI Conference on Human Factors in Computing Systems, pages 2782–2791. ACM, 2017.

[146] Jonathan Ledlie. Huzzah for my thing: Evaluating a pilot of a mobile service in kenya. *Qual Meets Quant, London, UK*, 2010.

[147] Hung-Chi Lee, Ya Fang Cheng, Szu Yang Cho, Hsien-Hui Tang, Jane Hsu, and Chien-Hsiung Chen. Picgo: Designing reminiscence and storytelling for the elderly with photo annotation. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems*, DIS Companion '14, pages 9–12, New York, NY, USA, 2014. ACM.

[148] Ann Light. HCI as heterodoxy: Technologies of identity and the queering of interaction with computers. *Interacting with Computers*, 23(5):430–438, 9 2011.

[149] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[150] Catherine Lim, Andrew B.L. Berry, Tad Hirsch, Andrea L. Hartzler, Edward H. Wagner, Evette Ludman, and James D. Ralston. "it just seems outside my health": How patients with chronic conditions perceive communication boundaries with providers. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, pages 1172–1184, New York, NY, USA, 2016. ACM.

[151] Joseph Lindley and Paul Coulton. Back to the future: 10 years of design fiction. *ACM International Conference Proceeding Series*, pages 210–211, 2015.

[152] Joseph Lindley and Paul Coulton. Pushing the limits of design fiction: The case for fictional research papers. *Conference on Human Factors in Computing Systems - Proceedings*, pages 4032–4043, 2016.

[153] Edwin A Locke and Gary P Latham. The development of goal setting theory: A half century retrospective. *Motivation Science*, 5(2):93, 2019.

[154] Abby M Lohr, Maia Ingram, Annabelle V Nuñez, Kerstin M Reinschmidt, and Scott C Carvajal. Community–clinical linkages with community health workers in the united states: A scoping review. *Health promotion practice*, 19(3):349–360, 2018.

[155] Steve Love and Mark Perry. Dealing with mobile conversations in public places: some implications for the design of socially intrusive technologies. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pages 1195–1198. ACM, 2004.

[156] Xiao Ma, Jeff Hancock, and Mor Naaman. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3857–3869. ACM, 2016.

[157] Xiao Ma, Jeffery T Hancock, Kenneth Lim Mingjie, and Mor Naaman. Self-disclosure and perceived trustworthiness of airbnb host profiles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2397–2409. ACM, 2017.

[158] Elizabeth Madigan. People with heart failure and home health care resource use and outcomes. *Journal of clinical nursing*, 17(7b):253–259, 2008.

[159] Meethu Malu, Nikunj Jethi, and Dan Cosley. Encouraging personal storytelling by example. In *Proceedings of the 2012 iConference*, pages 611–612. ACM, 2012.

[160] Anandi Mani, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. Poverty impedes cognitive function. *science*, 341(6149):976–980, 2013.

[161] Ahmad S. Manshad, Muhanad S. Manshad, and Sara S. Manshad. Real-time activity-sensitive wearable ankle edema monitoring system for elderly and visually impaired heart failure patients. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17, pages 391–392, New York, NY, USA, 2017. ACM.

[162] Elizabeth B. Marquis, Sangmi Kim, Rasha Alahmad, Casey S. Pierce, and Lionel P. Robert. Impacts of perceived behavior control and emotional labor on gig workers. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 241–244. Association for Computing Machinery, 10 2018.

[163] David Martin, Benjamin V. Hanrahan, Jacki O'Neill, and Neha Gupta. Being a turker. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14, pages 224–235, New York, NY, USA, 2014. ACM.

[164] Nancy Matthew-Maich, Lauren Harris, Jenny Ploeg, Maureen Markle-Reid, Ruta Valaitis, Sarah Ibrahim, Amiram Gafni, and Sandra Isaacs. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. *JMIR mHealth and uHealth*, 4(2), 2016.

[165] Indrani Medhi, Aman Sagar, and Kentaro Toyama. Text-free user interfaces for illiterate and semi-literate users. In *Information and Communication Technologies and Development, 2006. ICTD'06. International Conference on*, pages 72–82. IEEE, 2006.

[166] Indrani Medhi-Thies, Pedro Ferreira, Nakull Gupta, Jacki O'Neill, and Edward Cutrell. Krishipustak: a social networking system for low-literate

farmers. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1670–1681. ACM, 2015.

[167] Mikhail I Melnik and James Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. *The journal of industrial economics*, 50(3):337–349, 2002.

[168] Tamir Mendel and Eran Toch. Susceptibility to social influence of privacy behaviors: Peer versus authoritative sources. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 581–593. ACM, 2017.

[169] Medic Mobile. Medic mobile, 2018. Retrieved June 25, 2018 from `https://medicmobile.org/`.

[170] Shama Mohammed, Rachel Glennerster, and Aamir J Khan. Impact of a daily sms medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. *PloS one*, 11(11):e0162944, 2016.

[171] Anjali K Mohan, Edward Cutrell, and Balaji Parthasarathy. Instituting credibility, accountability and transparency in local service delivery?: helpline and aasthi in karnataka, india. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers-Volume 1*, pages 238–247. ACM, 2013.

[172] Maletsabisa Molapo, Melissa Densmore, and Brian DeRenzi. Video consumption patterns for first time smartphone users: Community health workers in lesotho. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6159–6170. ACM, 2017.

[173] Maletsabisa Molapo, Melissa Densmore, and Limpho Morie. Apps and skits: Enabling new forms of village-to-clinic feedback for rural health

education. In *Proceedings of the 7th Annual Symposium on Computing for Development*, page 10. ACM, 2016.

[174] Stephanie A Morey, Laura H Barg-Walkow, and Wendy A Rogers. Managing heart failure on the go: Usability issues with mhealth apps for older adults. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 1–5. SAGE Publications Sage CA: Los Angeles, CA, 2017.

[175] Sendhil Mullainathan. Sendhil mullainathan, 2019. Retrieved November 16, 2019 from `https://www.chicagobooth.edu/faculty/directory/m/sendhil-mullainathan/`.

[176] Sendhil Mullainathan and Richard H Thaler. Behavioral economics. Technical report, National Bureau of Economic Research, 2000.

[177] David Nemer, Ian Spangler, and Michaelanne Dye. Airbnb and the costs of emotional labor in havana, cuba. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 245–248. ACM, 2018.

[178] Francisco Nunes, Nervo Verdezoto, Geraldine Fitzpatrick, Morten Kyng, Erik Grönvall, and Cristiano Storni. Self-care technologies in hci: Trends, tensions, and opportunities. *ACM Trans. Comput.-Hum. Interact.*, 22(6):33:1–33:45, December 2015.

[179] Ted O'Donoghue and Matthew Rabin. Doing it now or later. *American Economic Review*, 89(1):103–124, 1999.

[180] Aisling Ann O'Kane, Yvonne Rogers, and Ann E. Blandford. Concealing or revealing mobile medical devices?: Designing for onstage and offstage presentation. In *Proceedings of the 33rd Annual ACM Conference on Human*

Factors in Computing Systems, CHI '15, pages 1689–1698, New York, NY, USA, 2015. ACM.

[181] Fabian Okeke, Lucas Nene, Anne Muthee, Stephen Odindo, Dianna Kane, Isaac Holeman, and Nicola Dell. Opportunities and challenges in connecting care recipients to the community health feedback loop. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, page 13. ACM, 2019.

[182] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. Good vibrations: can a digital nudge reduce digital overload? In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 4. ACM, 2018.

[183] Fabian Okeke, Michael Sobolev, and Deborah Estrin. Towards a framework for mobile behavior change research. In *Proceedings of the Technology, Mind, and Society*, page 27. ACM, 2018.

[184] Fabian Okeke, Emily Tseng, Benedetta Piantella, Mikaela Brown, Harveen Kaur, Madeline R Sterling, and Nicola Dell. Technology, home health care, and heart failure: a qualitative analysis with multiple stakeholders. In *Proceedings of the Conference on Computing & Sustainable Societies*, pages 122–133. ACM, 2019.

[185] Fabian Okeke, Beatrice Wasunna, Mercy Amulele, Isaac Holeman, and Nicola Dell. Including the voice of care recipients in community health feedback loops in rural kenya. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):71, 2019.

[186] Martin T Orne. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist*, 17(11):776, 1962.

[187] Joyojeet Pal, Anandhi Viswanathan, Priyank Chandra, Anisha Nazareth, Vaishnav Kameswaran, Hariharan Subramonyam, Aditya Johri, Mark S Ackerman, and Sile O'Modhrain. Agency in assistive technology adoption: Visual impairment and smartphone use in bangalore. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5929–5940. ACM, 2017.

[188] Somani Patnaik, Emma Brunskill, and William Thies. Evaluating the accuracy of data collection on mobile phones: A study of forms, sms, and voice. In *2009 International Conference on Information and Communication Technologies and Development (ICTD)*, pages 74–84. IEEE, 2009.

[189] Jennifer Pearson, Simon Robinson, Matt Jones, Amit Nanavati, and Nitendra Rajput. Acqr: acoustic quick response codes for content sharing on low end phones with no internet connectivity. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 308–317. ACM, 2013.

[190] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[191] Lucy Pei and Bonnie Nardi. We did it right, but it was still wrong: Toward assets-based design. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, page alt07. ACM, 2019.

[192] Trevor Perrier, Nicola Dell, Brian DeRenzi, Richard Anderson, John Kinuthia, Jennifer Unger, and Grace John-Stewart. Engaging pregnant women in kenya with a hybrid computer-human sms communication system. In

Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 1429–1438. ACM, 2015.

[193] Trevor Perrier, Brian DeRenzi, and Richard Anderson. Ussd: the third universal app. In *Proceedings of the 2015 Annual Symposium on Computing for Development*, pages 13–21. ACM, 2015.

[194] Trevor Perrier, Elizabeth K Harrington, Keshet Ronen, Daniel Matemo, John Kinuthia, Grace John-Stewart, Richard Anderson, and Jennifer A Unger. Male partner engagement in family planning sms conversations at kenyan health clinics. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, page 3. ACM, 2018.

[195] Fahad Pervaiz, Trevor Perrier, Sompasong Phongphila, and Richard Anderson. User errors in sms based reporting systems. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, page 55. ACM, 2015.

[196] PHI. U.s. home care workers: Key facts, 2018. Retrieved Dec 24, 2018 from `https://phinational.org/wp-content/uploads/legacy/phi-home-care-workers-key-facts.pdf`.

[197] Rowan Phipps, Shrirang Mare, Peter Ney, Jennifer Webster, and Kurtis Heimerl. Thinsim-based attacks on mobile money systems. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, page 23. ACM, 2018.

[198] Anthony Poon, Sarah Giroux, Parfait Eloundou-Enyegue, François Guimbretière, and Nicola Dell. Engaging high school students in cameroon with exam practice quizzes via sms and whatsapp. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 482. ACM, 2019.

[199] Shantanu Prabhat, Sneha Nanavati, and Nimmi Rangaswamy. India's "uberwallah": Profiling Uber drivers in the gig economy. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 1 2019.

[200] Damaraju Raghavarao and Walter T Federer. Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):40–45, 1979.

[201] Kiran Gopakumar Rajalekshmi. E-governance services through telecenters: The role of human intermediary and issues of trust. *Information Technologies & International Development*, 4(1):pp–19, 2007.

[202] Divya Ramachandran, John Canny, Prabhu Dutta Das, and Edward Cutrell. Mobile-izing health workers in rural india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1889–1898. ACM, 2010.

[203] Noopur Raval and Paul Dourish. Standing out from the crowd: Emotional labor, body labor, and temporal labor in ridesharing. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, volume 27, pages 97–107, 2016.

[204] Subhajyoti Ray. Reinforcing accountability in public services: an ict enabled framework. *Transforming Government: People, Process and Policy*, 6(2):135–148, 2012.

[205] Fred Reichheld. The ultimate question. *Harvard Business School Press, Boston, MA*, 2006.

[206] Frederick F Reichheld and Rob Markey. *The Ultimate Question 2.0: How net promoter companies thrive in a customer-driven world*. Harvard business press, 2011.

[207] David Russell, Ana Mola, Nicole Onorato, Sheniqua Johnson, Jessica Williams, Mark Andaya, and Marki Flannery. Preparing home health aides to serve as health coaches for home care patients with chronic illness: Findings and lessons learned from a mixed-method evaluation of two pilot programs. *Home Health Care Management & Practice*, 29(3):191–198, 2017.

[208] Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. Intermediated technology use in developing communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2583–2592. ACM, 2010.

[209] Nithya Sambasivan and Edward Cutrell. Understanding negotiation in airtime sharing in low-income microenterprises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 791–800. ACM, 2012.

[210] Alain Samson. The behavioral economics guide 2015 (with an introduction by dan ariely), 2015.

[211] David Sando, Hannah Ratcliffe, Kathleen McDonald, Donna Spiegelman, Goodluck Lyatuu, Mary Mwanyika-Sando, Faida Emil, Mary Nell Wegner, Guerino Chalamilla, and Ana Langer. The prevalence of disrespect and abuse during facility-based childbirth in urban tanzania. *BMC pregnancy and childbirth*, 16(1):236, 2016.

[212] Mark Di Sano, Andres Perez, Miguel A. Labrador, Ponrathi Athilingam, and Federico Giovannetti. Heartmapp: A mobile application to improve chf outcomes and reduce hospital readmissions. In *Proceedings of the Conference on Wireless Health*, WH '15, pages 20:1–20:7, New York, NY, USA, 2015. ACM.

[213] Marén Schorch, Lin Wan, David William Randall, and Volker Wulf. Designing for those who are overlooked: Insider perspectives on care practices and cooperative work of elderly informal caregivers. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 787–799, New York, NY, USA, 2016. ACM.

[214] Janet Schwartz, Daniel Mochon, Lauren Wyper, Josiase Maroba, Deepak Patel, and Dan Ariely. Healthier by precommitment. *Psychological science*, 25(2):538–546, 2014.

[215] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. Reflective design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility*, CC '05, pages 49–58, New York, NY, USA, 2005. ACM.

[216] Eldar Shafir. Eldar shafir, 2019. Retrieved November 16, 2019 from `https://www.princeton.edu/haushofer/`.

[217] Muzafer Sherif. A study of some social factors in perception. *Archives of Psychology (Columbia University)*, 1935.

[218] Anita Shet, Ayesha De Costa, N Kumarasamy, Rashmi Rodrigues, Bharat Bhusan Rewari, Per Ashorn, Bo Eriksson, and Vinod Diwan. Effect of mobile telephone reminders on treatment outcome in hiv: evidence from a randomised controlled trial in india. *Bmj*, 349:g5978, 2014.

[219] Emrys Shoemaker, Gudrun Svava Kristinsdottir, Tanuj Ahuja, Dina Baslan, Bryan Pon, Paul Currion, Pius Gumisizira, and Nicola Dell. Identity at the margins: Examining refugee experiences with digital identity systems in lebanon, jordan, and uganda. 2019.

[220] Debra Singh, Joel Negin, Michael Otim, Christopher Garimoi Orach, and Robert Cumming. The effect of payment and incentives on motivation

and focus of community health workers: five case studies from low-and middle-income countries. *Human resources for health*, 13(1):58, 2015.

[221] Burrhus Frederic Skinner. *Science and human behavior*. Number 92904. Simon and Schuster, 1965.

[222] Burrhus Frederic Skinner. *The behavior of organisms: An experimental analysis*. BF Skinner Foundation, 1990.

[223] Kiley Sobel, Geza Kovacs, Galen McQuillen, Andrew Cross, Nirupama Chandrasekaran, Nathalie Henry Riche, Ed Cutrell, and Meredith Ringel Morris. Edufeed: A social feed to engage preliterate children in educational activities. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 491–504. ACM, 2017.

[224] Hannah Spring, Saugato Datta, and Sabitri Sapkota. Using behavioral science to design a peer comparison intervention for postabortion family planning in nepal. *Frontiers in public health*, 4:123, 2016.

[225] Vivek Srinivasan, Vibhore Vardhan, Snigdha Kar, Siddhartha Asthana, Rajendran Narayanan, Pushpendra Singh, Dipanjan Chakraborty, Amarjeet Singh, and Aaditeshwar Seth. Airavat: An automated system to increase transparency and accountability in social welfare schemes in india. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2*, pages 151–154. ACM, 2013.

[226] Thomas K Srull and Robert S Wyer. The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social psychology*, 37(10):1660, 1979.

[227] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. Beyond self-tracking and reminders: designing smartphone apps that support habit forma-

tion. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 2653–2662. ACM, 2015.

[228] Bruce Sterling. Shaping Things. *October*, 39:152, 2005.

[229] Madeline R Sterling, Amy L Shaw, Peggy BK Leung, Monika M Safford, Christine D Jones, Emma K Tsui, and Diana Delgado. Home care workers in heart failure: a systematic review. *Journal of multidisciplinary healthcare*, 11:481, 2018.

[230] Madeline R Sterling, Ariel F Silva, Peggy BK Leung, Amy L Shaw, Emma K Tsui, Christine D Jones, Laura Robbins, Yanira Escamilla, Ann Lee, Faith Wiggins, Frances Sadler, Martin F Shapiro, Mary E Charlson, Lisa M Kern, and Monika M Safford. ”it’s like they forget that the word ’health’ is in ’home health aide”’: Understanding the perspectives of home care workers who care for adults with heart failure. *Journal of the American Heart Association*, 7(23):e010134, 2018.

[231] Madeline R Sterling, Ariel F Silva, Amy L Shaw, Peggy B Leung, Emma K Tsui, Christine D Jones, Laura Robbins, Ann Lee, Yanira Escamilla, Frances Sadler, et al. Home care workers who care for adults with heart failure. *Circulation: Cardiovascular Quality and Outcomes*, 11(Suppl_1):A16, 2018.

[232] Robyn Stone, Janet P Sutton, Natasha Bryant, Annelise Adams, and Marie Squillace. The home health workforce: A distinction between worker categories. *Home health care services quarterly*, 32(4):218–233, 2013.

[233] Robyn I Stone. The direct care worker: The third rail of home care policy. *Annu. Rev. Public Health*, 25:521–537, 2004.

[234] Anselm Strauss and Juliet M Corbin. *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Publications, Inc, 1990.

[235] Victor J Strecher, Gerard H Seijts, Gerjo J Kok, Gary P Latham, Russell Glasgow, Brenda DeVellis, Ree M Meertens, and David W Bulger. Goal setting as a strategy for health behavior change. *Health education quarterly*, 22(2):190–200, 1995.

[236] Sharifa Sultana, François Guimbretière, Phoebe Sengers, and Nicola Dell. Design within a patriarchal society: Opportunities and challenges in designing for rural women in bangladesh. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 536. ACM, 2018.

[237] Steven Tadelis. Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8:321–340, 2016.

[238] Africa's Talking. Africa's talking, 2019. Retrieved March 26, 2019 from `https://africastalking.com/`.

[239] Nick Taylor, Keith Cheverst, Peter Wright, and Patrick Olivier. Leaving the wild: lessons from community technology handovers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1549–1558. ACM, 2013.

[240] Richard H Thaler and Cass R Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[241] Jennifer Thom, David Millen, and Joan DiMicco. Removing gamification from an enterprise sns. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 1067–1070. ACM, 2012.

[242] Shari Trewin, Diogo Marques, and Tiago Guerreiro. Usage of subjective scales in accessibility research. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 59–67. ACM, 2015.

[243] Emily Tseng, Fabian Okeke, Madeline R Sterling, and Nicola Dell. "we can learn. why not?": Challenges and opportunities in designing for equity for home health aides. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 122–133. ACM, In submission, 2019.

[244] Anupriya Tuli, Shruti Dalvi, Neha Kumar, and Pushpendra Singh. "it's a girl thing": Examining challenges and opportunities around menstrual health education in india. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(5):29, 2019.

[245] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[246] U-Report, 2019. Retrieved March 21, 2019 from `http://www.ureport.in/`.

[247] JA Unger, Keshet Ronen, Trevor Perrier, Brian DeRenzi, Jennifer Slyker, AL Drake, Dyphna Mogaka, John Kinuthia, and Grace John-Stewart. Short message service communication improves exclusive breastfeeding and early postpartum contraception in a low-to middle-income country setting: a randomised trial. *BJOG: An International Journal of Obstetrics & Gynaecology*, 125(12):1620–1629, 2018.

[248] Anna Vallgrda, Erik Grönvall, and Jonas Fritsch. Experimental engineering: Articulating and valuing design experimentation. In *DIS 2017 Companion - Proceedings of the 2017 ACM Conference on Designing Interactive Systems*, pages 197–201. Association for Computing Machinery, Inc, 6 2017.

[249] Edwin AJ van Hooft, Marise Ph Born, Toon W Taris, Henk Van der Flier, and Roland WB Blonk. Bridging the gap between intentions and behavior: Implementation intentions, action control, and procrastination. *Journal of vocational behavior*, 66(2):238–256, 2005.

[250] Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. Sangeet Swara: A Community-Moderated Voice Forum in Rural India. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 417–426, New York, NY, USA, 2015. ACM.

[251] Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. Social media platforms for low-income blind people in india. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 259–272. ACM, 2015.

[252] Aditya Vashistha, Edward Cutrell, and William Thies. Increasing the reach of snowball sampling: The impact of fixed versus lottery incentives. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1359–1363. ACM, 2015.

[253] Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. Mobile Video Dissemination for Community Health. In *Proceedings of ICTD*, 2016.

[254] Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. Examining localization approaches for community health. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pages 357–368. ACM, 2017.

[255] Aditya Vashistha, Fabian Okeke, Richard Anderson, and Nicola Dell. "you can always do better!" the impact of social proof on participant response bias. 2018.

[256] Aditya Vashistha and William Thies. Ivr junction: Building scalable and distributed voice forums in the developing world. In *NSDR*, 2012.

[257] Meera Viswanathan, Jennifer L Kraschnewski, Brett Nishikawa, Laura C Morgan, Amanda A Honeycutt, Patricia Thieda, Kathleen N Lohr, and Daniel E Jonas. Outcomes and costs of community health worker interventions: a systematic review. *Medical care*, pages 792–808, 2010.

[258] Michalis Vitos, Julia Altenbuchner, Matthias Stevens, Gillian Conquest, Jerome Lewis, and Muki Haklay. Supporting collaboration with non-literate forest communities in the congo-basin. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1576–1590. ACM, 2017.

[259] S. L. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–66, March 1965.

[260] Galen Weld, Trevor Perrier, Jenny Aker, Joshua E. Blumenstock, Brian Dillon, Adalbertus Kamanzi, Editha Kokushubira, Jennifer Webster, and Richard J. Anderson. ekichabi: Information access through basic mobile phones in rural tanzania. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 133:1–133:12, New York, NY, USA, 2018. ACM.

[261] David Wendler and Franklin G Miller. Deception in the pursuit of science. *Archives of Internal Medicine*, 164(6):597–600, 2004.

[262] Caroline Whidden, Kassoum Kayentao, Jenny X Liu, Scott Lee, Youssouf Keita, Djoumé Diakité, Alexander Keita, Samba Diarra, Jacqueline Edwards, Amanda Yembrick, et al. Improving community health worker performance by using a personalised feedback dashboard for supervision: a randomised controlled trial. *Journal of global health*, 8(2), 2018.

[263] Kevin Winoto, Piao Chen, Yongfu Wang, and Pinata Winoto. Metap: A personalized spatial memory training for improving functional autonomy of adults with mild cognitive impairments in china. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, DIS '18 Companion, pages 249–252, New York, NY, USA, 2018. ACM.

[264] Niall Winters and Kentaro Toyama. Human-computer interaction for development: mapping the terrain. *Information Technologies & International Development*, 5(4), 2009.

[265] Fang Wu and Bernardo Huberman. How public opinion forms. *Internet and Network Economics*, pages 334–341, 2008.

[266] Susan P Wyche, Sarita Yardi Schoenebeck, and Andrea Forte. Facebook is a luxury: An exploratory study of social media use in rural kenya. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 33–44. ACM, 2013.

[267] Naomi Yamashita, Hideaki Kuzuoka, Keiji Hirata, Takashi Kudo, Eiji Aramaki, and Kazuki Hattori. Changing moods: How manual tracking by family caregivers improves caring and family communication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 158–169, New York, NY, USA, 2017. ACM.

[268] Naomi Yamashita, Hideaki Kuzuoka, Takashi Kudo, Keiji Hirata, Eiji Aramaki, and Kazuki Hattori. How information sharing about care recipients by family caregivers impacts family communication. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 222:1–222:13, New York, NY, USA, 2018. ACM.