From User-Centered to Adoption-Centered Design: A Case Study of an HCI Research Innovation Becoming a Product

Parmit K. Chilana[1], Amy J. Ko[2] and Jacob O. Wobbrock[2]

[1]Management Sciences
University of Waterloo
Waterloo, ON Canada N2L 3G1
pchilana@uwaterloo.ca

[2]The Information School | DUB Group
University of Washington
Seattle, WA 98195 USA
{ajko, wobbrock}@uw.edu

ABSTRACT

As we increasingly strive for scientific rigor and generalizability in HCI research, should we entertain any hope that by doing good science, our discoveries will eventually be more transferrable to industry? We present an in-depth case study of how an HCI research innovation goes through the process of transitioning from a university project to a revenue-generating startup financed by venture capital. The innovation is a novel contextual help system for the Web, and we reflect on the different methods used to evaluate it and how research insights endure attempted dissemination as a commercial product. Although the extent to which any innovation succeeds commercially depends on a number of factors like market forces, we found that our HCI innovation with user-centered origins was in a unique position to gain traction with customers and garner buy-in from investors. However, since end users were not the buyers of our product, a strong user-centered focus obfuscated other critical needs of the startup and pushed out perspectives of non-user-centered stakeholders. To make the research-to-product transition, we had to focus on *adoption-centered design*, the process of understanding and designing for adopters and stakeholders of the product. Our case study raises questions about how we evaluate the novelty and research contributions of HCI innovations with respect to their potential for commercial impact.

Keywords

Commercialization, productization, dissemination, research impact, technology transfer, adoption-centered design.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (*e.g.*, HCI): User Interfaces—*evaluation / methodology*.

INTRODUCTION

What constitutes success for a system innovation emerging from human-computer interaction (HCI) research? One way to address this question might be to look at how well the design is validated in the target context and to what extent its findings are generalizable. Yet there are numerous debates and disagreements within HCI about what constitutes an adequate systems evaluation [8,22]. Some scholars have even claimed that the bar for evaluation at UIST, CHI, and other HCI publication venues is increasingly unrealistic for systems papers [1,14] by insisting on having results from controlled studies or large cohorts of representative users.

Since HCI is still an evolving field that is striving to bridge multiple disciplines, debates and disagreements about adequate system evaluation and what makes a system successful are necessary. Part of doing "good science" is ensuring that our work has rigor and that we are able to match our claimed contributions with actual findings. But as we strive for scientific rigor and generalizability in research, should we entertain any hope that by doing good science, our discoveries will eventually be more transferrable to users or customers? Or are such concerns beyond the scope of applied technology researchers in HCI? In short, must we care?

Despite decades of HCI research, these are not easy questions to answer and indeed, they have rarely been tackled head-on (with a few noteworthy exceptions [7,9,11]). There has been repeated concern that even though we are accumulating great innovations in the HCI field, we rarely see HCI innovations develop into commercial products [11]. Given that researchers usually lack the knowledge, resources, connections, experience, interest, or time to pursue commercialization [26], perhaps the lack of products emerging from HCI research is not surprising. But could another hindrance to HCI technology transfer be that we often look at our innovations "too narrowly" by focusing on generalizability only from end users' perspectives [9]? What if end users are not the paying customers so vital for getting a product to market?

For any commercial product to successfully emerge, there are multiple stakeholders beyond end users, such as business customers, target buyers, approvers, administrators, regulators, financers, reporters, analysts, and more [5]. Although we have seen the development of methods in HCI and usability practice that allow researchers and practitioners to understand stakeholder needs [23], we rarely see these methods employed in system evaluations for the sake of beyond-the-user market adoption. Could these additional perspectives give HCI innovations better chances at being adopted? Or would employing such methods distract from delivering high-quality, innovative research?

CHI 2015, April 18-23, 2015, Seoul, Republic of Korea

http://dx.doi.org/10.1145/2702123.2702412

In this paper, we investigate how one HCI innovation is transitioning from a research project to a commercial product and startup business, and how well the research foundation for the innovation generalizes along this path. Specifically, we present an in-depth case study of several evaluation methods applied to *LemonAid* [4] (Figure 1) over the last five years, from its inception as a university research project to commercialization through a venture capital financed startup called *AnswerDash*.[1] We trace the motivations for adopting different HCI methods at different stages during the evolution of the research, product, and startup business and the tradeoffs made between user-centered design and what we call *adoption-centered design*—understanding and designing for product adopters and stakeholders beyond just end users.

The main contribution of this paper is providing insight into how an HCI innovation can evolve from an academic endeavor to a commercial product when the end user is not the main adopter or customer of the system. In our discussion, we tackle the question of whether HCI technology researchers should develop and employ methods for concerning themselves with beyond-the-user stakeholder concerns—stakeholders that lie along the "adoption path" that must be trod for innovations to gain market traction and become widespread.

RELATED WORK

To contextualize our case study, we focused our literature review on three areas: (1) perspectives on evaluation and generalizability of HCI innovations; (2) experiences of technology transfer from research in related areas, such as software engineering; and (3) the challenges inherent in the process of commercializing disruptive innovations.

Generalizability and System Evaluation in HCI Research

The concept of generalizability in research refers to the extent to which we can use research results based on specific cases to form abstractions about the general case, or the degree to which empirical results can be generalized to an entire population [16,17]. In the context of HCI, an important form of generalization is the extent to which the design of a system is applicable to a broader set of target users [18]. In the design stage, we strive for generalizability by moving from individual user data (*i.e.*, through formative studies) to general requirements for the system being designed. In the evaluation stage, generalization becomes the process of determining the extent to which study results achieved with a particular population doing specific tasks with a system can apply to the target user population [18].

The extent to which we can demonstrate generalizability varies based on the research domain and type of method used [16]. In technical HCI research, the debate has often been about showing the generalizability of interaction techniques versus whole interactive systems. For example, unlike novel interaction techniques that can be validated through controlled experimentation, novel systems can be difficult to evaluate as they often introduce new platforms or capabilities that do not have obvious benchmark systems for comparison [8,22]. Systems in emerging areas such as social computing face additional intractable challenges—for example, evaluations may be penalized for snowball sampling or for not demonstrating exponential growth [1].

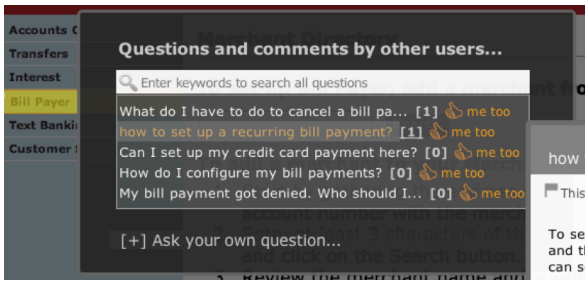

Figure 1. The LemonAid retrieval interface.

Increasingly, evaluations of systems in HCI stress external validity and ecological validity for increased generalizability [18]. Citing the potential perils of limited usability tests, questions have risen about prematurely validating an innovation before "a culture" is formed around it [8]. Similarly, emphasis is being placed on the need for evaluations to consider more realistic situations, tasks, and users [22].

At what point do we know that our evaluation is sufficient to demonstrate success, at least in research? What if end users are not the eventual customers, adopters, or buyers of our HCI innovations? Does the soundness of our user-centered methods matter to transferability? Although ideas about the applicability and generalizability of HCI research systems have been previously raised, to our knowledge, no work has directly investigated the link between the soundness of HCI technology research conducted and the likelihood that a system will eventually see widespread adoption and market traction. Our case study adds another dimension to this debate: if we care about adoption, and if adoption depends on more than end users, are user-centered evaluation criteria enough to maximize our chances for success?

Perspectives on Tech-Transfer in Software Engineering

While HCI is a relatively young field, technology transfer has a long history in more established domains, such as software engineering. Experiences of software engineering researchers going through technology transfer offer a wide range of perspectives—from internal transfer at NASA [27] and IBM [6], to formation of a startup from a research lab [19], to several other examples of tech-transfer successes and failures in software engineering research [15,26]. One of the main themes reflected in these perspectives is that successful tech-transfer is not only about an idea or its implementation—it is about knowing the audience for the technology and establishing relationships with customers to understand market needs.

Our experience shows that compared to less user-centered computing innovations, HCI innovations can perhaps get a "head start" on adoption because they inherently consider end users as the audience. However, similar to questions

[1] http://www.answerdash.com

that have already been raised about HCI technology transfer [7,9,11], we also ask if the strong focus on end users is necessary for having impact through commercialization if software ventures clearly have different goals.

Disruptive Innovations in the Marketplace

While our case study raises some questions about the limits of user-centered evaluation methods for achieving commercialization, we realize that it is also important to understand perspectives on market forces that are not unique to HCI, but bound to affect *any* innovation [21].

For example, Christenson [5] coined the term, *disruptive innovation* to describe a technology that shakes up an existing market because of the radical change it brings, but may, in fact, be "too radical." Wary of solutions ahead of their time, market leaders are often inclined to support *sustaining innovations*, which are well-matched to well-understood customer needs. Although disruptive innovations can be successful if enough early adopters are convinced and an appropriate market strategy can be created, commercialization of such innovations is a challenge [24].

Our case study shows that even though our novel contextual help tool succeeded in several user-centered research evaluations, it faced challenges while entering the mainstream market through early adopters. Such challenges were not anticipated during the research phase where the focus was on end user perceptions and usage.

BACKGROUND OF THE INNOVATION

The case study presented here is based on LemonAid, a novel selection-based crowdsourced contextual help tool and its transition to AnswerDash, a startup company based on LemonAid. We first give an overview of the original HCI research and its key features. For detailed information on LemonAid, readers are directed to our previous papers [3,4].

Motivation for LemonAid

Our initial motivation for inventing LemonAid was to help end users find answers to their questions that arise while using websites. Web-based technical support such as discussion forums, knowledge bases, FAQ pages, and social networking sites have been successful at ensuring that most technical support questions eventually receive helpful answers. Unfortunately, *finding* these answers is still difficult, since the queries that end users submit to search engines are often incomplete, imprecise, or use different vocabularies to describe the same problem. LemonAid improved the retrieval of relevant help by introducing a selection-based contextual help approach that embedded end users' questions and answers directly in the user interface (UI).

LemonAid User Interface and System Design

Instead of searching the web or scouring web forums, lists of frequently asked questions, or other knowledge bases, LemonAid allowed users to retrieve help by selecting a label, widget, image, or other UI element on any page of a website. The user's selection of a UI element triggered the retrieval of questions related to the selection (Figure 1) by matching the selection's contextual data against all the questions in the repository. The contextual data consisted of the underlying Document-Object Model (DOM) element's text, tag type, and location within the UI (described in detail in [3]). The retrieval algorithm used a relevance ranking approach, sorting questions based on the highest-scoring matches. In addition to viewing answers, users could also search for another question, submit their own question, or vote on the helpfulness of answers.

LemonAid worked with standard HTML pages, without making any assumptions about the implementation of the underlying website and did not require developers to use any specific UI toolkits. LemonAid presented contextual Q&A in a separate visual layer atop an application's UI that did not interfere with the underlying site's functionality. To add LemonAid to a site, developers had to extract text labels appearing in the UI from a web application's code and include the LemonAid framework in their client-side deployment with one line of JavaScript. The vision was that this small amount of technical work would significantly lower the barrier to adoption.

TOWARDS A GENERALIZABLE USER-CENTERED DESIGN: EVOLUTION OF THE RESEARCH PROTOTYPE

Consistent with expectations for an academic HCI research project, we focused on addressing user-centered concerns and showing the validity of our innovation from the end user's perspective. We first carried out a formative evaluation to inform the system and algorithm design, followed by an evaluation of the system's feasibility using crowdsourcing, and finally, a longitudinal field study to assess the innovation's ecological validity. As we progressed through each of these stages, our goal was to explicitly increase the scope of our innovation's generalizability (Figure 2). However, these activities were not conducted with commercialization in view; they were strictly knowledge-generating research activities. In later sections of this paper, we describe our commercialization efforts and how we had to make the transition to focus more on adopter and stakeholder perspectives rather than end users.

How do we solve the general problem? Informing the system design with a user study

The core problem that we first tackled was designing a way for end users to retrieve relevant questions asked by other end users on the website. The goal was to find a design so-

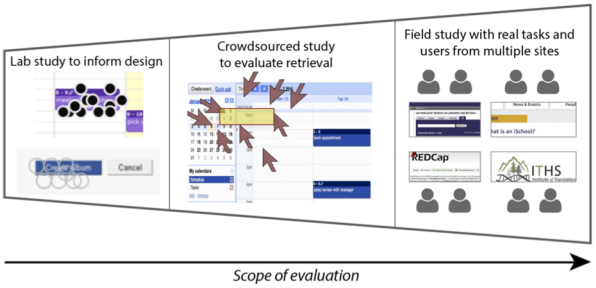

Figure 2. Increasing scope of evaluation at different stages to inform and evaluate design.

lution that would generalize across different sites and help scenarios and not be tied to the underlying implementation of the website or web application.

In exploring the design space of this concept, there were different types of contextual data that LemonAid could gather from the website to associate with relevant Q&A. Since building and testing all of these possible solutions would be time-intensive and costly, we decided to first use a low-fidelity paper-based approach to discern which aspects of the interface might be useful in discriminating between different help problems in the application.

We designed a paper-based *Wizard-of-Oz* study that reflected 12 different help-seeking scenarios from the help forums of Facebook, Google, and Wikipedia. The paper showed participants a printed screenshot and a brief paraphrased description of a help need. We provided physical stickers and asked participants to pretend that they had a "magic wand" to point anywhere in the interface to get help, indicated by placing the sticker somewhere on the screenshot.

This low-fidelity approach provided us with the key insight for designing a generalizable version of our selection-based retrieval interaction and algorithm: participants tended to select similar application-specific UI elements for similar help needs and different UI elements for different help needs, making UI elements useful discriminators. These findings suggested that LemonAid could determine similarity between selections largely based on the text on UI elements and leverage additional attributes of the selected DOM object such as its layout position and appearance.

Over the last five years, this key insight about how users can retrieve relevant Q&A by making UI element selections has been validated repeatedly. However, at the time, our paper-based study could only inform the novel interaction design, not provide evidence about the effectiveness of the retrieval approach for real-world sites and applications.

Does the system behave as intended? Assessing system feasibility using simulated data

In our second evaluation, our focus shifted to investigating the technical feasibility and effectiveness of our selection-based contextual retrieval interaction and algorithm. In particular, we focused on answering the following research question: *across a corpus of help-problem scenarios, how effective is LemonAid at retrieving relevant Q&A asked by other users based only on the current user's selection*?

To address this question, we used the Mechanical Turk (mTurk) platform to get access to a large number of users and their help selections. Alternatives, such as simulating the selections or running a lab study, were not feasible as we wanted our corpus to reflect crowdsourced selections.

Similar to the protocol in our formative study, we generated detailed help scenarios created from a random sample of 50 questions in *Google Calendar's* forums and instrumented different interactive screens of this calendar application to collect users' selections. Over 500 users on mTurk read one or more of 50 help scenarios and performed a help request by selecting a UI element relevant to those help scenarios.

We ran our retrieval algorithm against the 2,748 selections generated by the mTurk study. Since LemonAid used a ranked retrieval approach, we assessed our algorithm's performance based on the rank of the first relevant question among all of the retrieved questions matching an end user's selection. LemonAid's performance was promising, retrieving 1 or more results for 90.3% of the selections with a median rank of 2 across the whole corpus, meaning the relevant result was in the top 2 results most of the time.

In addition to the overall performance of the retrieval algorithm, it was important for us to understand its performance over time, as users were likely to ask more questions. In an initial test of the effect of corpus size, we found that the median rank of the results degraded as the number of selections increased, but that this was a slow degradation. Still, determining whether LemonAid's algorithm would scale to any live application and a growing database of Q&A was not considered within the scope of our evaluation. Our focus was primarily on the acceptability to users from an interactive perspective and the quality of the Q&A retrieval.

How does the system behave in the "real world?" Assessing ecological validity with field deployments

Although our retrieval evaluation was promising, its promise was based on pre-defined tasks and a simulated community of users through mTurk, which did not reflect how users may actually use Q&A sites for troubleshooting. Thus, we did not know whether the retrieval approach we developed was ecologically valid or actually helpful for users' real tasks in real applications. To investigate these questions, we worked on deploying LemonAid "in the wild" [2].

Since LemonAid was not designed to be a tool that end users could download on their own, we had to first find website owners who would be willing to add LemonAid to their site. We began by contacting a number of web teams at our university who offered web applications to support teaching and research. From the start, we had to "sell" our solution and its benefits to would-be adopters. Over the course of several weeks, we had many meetings with over a dozen software developers and site owners across campus. We showed them demos of what LemonAid would look like on their site, clarified that they did not need to modify their site, and assured them that they would only have to commit minimal time to seeding the initial Q&A database and answer new questions. After each meeting, we iterated on our design and implemented additional features to facilitate adoption—for example, we built features for monitoring and flagging user content, allowing teams to receive notifications when new questions were asked, and building basic analytics to show the evolution of Q&A during the deployment period. Since some of these university-based sites had thousands or more visitors a day, we had to spend additional weeks engineering our prototype to be robust across our field sites.

The four sites that adopted LemonAid included the university library's main homepage, our academic department's site, a clinical data capture application, and a personnel and grant management site at the university's medical school. The deployments across the four sites lasted between 6-15 weeks. A common adoption motivation among the host teams was to move away from one-on-one assistance, which was difficult for them to maintain with small teams. They were intrigued by the idea of answering a question once and making its retrieval easier for future end users.

The teams that were reluctant to try LemonAid were mainly concerned about the time it might take from their already constrained schedules and the quality of the content that end users might contribute. They also hesitated to change the look and feel of their website, even for a few days during the deployment study. (As we will discuss in our next section, uncovering more of these adoption concerns would have been helpful for our later commercialization efforts, but during our research, we were only focused on showing the validity of our design from the end user's perspective.)

Our main research questions were assessing whether the LemonAid Q&A retrieval approach would be helpful, usable, and desirable for reuse from the perspective of end users. Using data triangulation [12], we assessed these variables from over 1200 usage logs, 168 exit survey responses, and 36 follow-up interviews. Overall, we found that end users responded positively to LemonAid across all of our deployments. Our analysis showed that they selected "helpful" on answers over 73 % of the time across the four sites and this finding corroborated with the helpfulness responses on the survey. Further analysis showed that new or infrequent site users found LemonAid to be more helpful than frequent users, particularly for learning about unfamiliar features. In addition, the survey responses across all deployments indicated that over 72% of end users agreed that LemonAid was intuitive and that they would use this help feature again.

These results on end users' perspectives provided a strong contrast to literature on help systems that show users often fear clicking on help after their initial exposure to a help system [20]. Our findings suggested there was high potential in helping users retrieve relevant help using our approach on different kinds of websites and applications.

After our deployments, we again probed into the software team perspective and interviewed team members who had integrated LemonAid onto their site. In particular, the teams were intrigued by LemonAid's basic analytics dashboard that aggregated frequently accessed questions and their locations of occurrence. The team members agreed that they were not able to obtain such analytics data from anywhere else and that it could be a useful way to augment their existing usage-based analytics services, like Google Analytics.

Although our field study showed that LemonAid was successful in helping end users locate relevant Q&A, it also shed light on other technical and organizational issues that would have to be addressed for users to actually benefit from this tool. For example, whether LemonAid would actually be available for end users would require a website owner to adopt it effectively. However, since such concerns were not directly relevant to assessing the novelty or the generalizability of our approach for end users, we did not investigate them in depth.

Summary of Research Evaluations

After these three evaluations, we had learned many things. First, we had reasonable confidence that our interaction design was usable and the underlying algorithm could effectively retrieve relevant Q&A. We also had heard feature requests and concerns that software teams had in adopting our solution and made improvements accordingly. Overall, the use of multiple evaluations of formative, technical, and ecological forms helped us to capture the end users' perspectives at different points in the design process.

Second, it became clear that there was a lot we did not know. Because our evaluations made inevitable tradeoffs between generalizability, precision, and realism [17], primarily in the size of our samples, it was difficult to say with confidence how our findings would fare beyond a university setting. In fact, while we were publishing our findings, we were also beginning conversations with investors, who naturally wondered whether our initial studies were enough evidence of success for financing, let alone whether paying customers would see business reasons to adopt something for which we had mainly demonstrated end user value.

In reflecting on our experience as researchers, there was little incentive for us to invest more time and resources during our field deployments in understanding adopters and other stakeholder issues. Such research may not strengthen the research contribution of our innovation; neither would it lead to more publications. Moreover, releasing our innovation online as an open source project would not likely lead to significant uptake. However, given the importance of these issues for commercialization, our field evaluations raised crucial questions about whether "potential for adoption" is something that can be assessed effectively when evaluating HCI research systems during peer review.

BEYOND END USERS: TRANSITIONING FROM A RESEARCH PROJECT TO A COMMERCIAL PRODUCT

With the benefits to end users arising from our field deployments, the rapid growth of e-commerce websites and web applications in the past decade, and the simplicity of implementing our contextual help solution, we felt there could be widespread value created by commercializing LemonAid. In this section, we highlight our transition from user-centered to adoption-centered design and the key questions we tackled in translating LemonAid into a commercially viable product now known as AnswerDash (Figure 3).

Startup Fundraising and Team Building

Using demos from our research prototype and findings from our field deployments, we were able to raise $50,000 from our university commercialization gap fund. The second and third authors used this funding in 2012-2013 to attract pilot

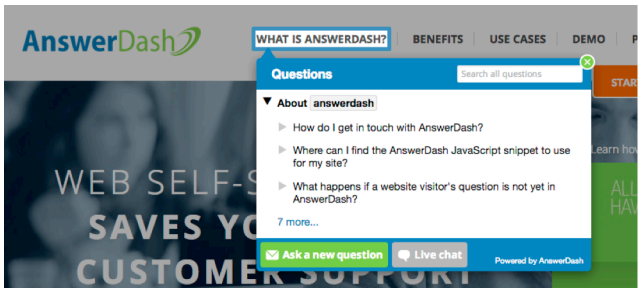

Figure 3: The AnswerDash end user interface showing questions retrieved on a UI element, in this case a link label.

customers, develop a beta product, secure the help of entrepreneurs-in-residence, and attract investors.

Our fundraising efforts revealed a number of concerns well outside the scope of our HCI research efforts. Investors rarely asked about our technology *per se*. Instead, they probed our understanding of the customers we were selling to, which type of customers saw the greatest benefit from the technology, how many companies might be viable customers, our strategies for acquiring customers through sales and marketing, and how we would build a top-notch go-to-market team. Although our HCI research taught us little about these concerns, our initial success at selling our solution to customers provided enough confidence to investors that we were able to raise a round of venture capital to launch a business. The purpose of the investment was for building the v1.0 product, developing a top-notch team, identifying a target market, and securing the initial 50+ customers within that market.

Now with funds, we began to hire. We began with engineers, as they would have to build the v1.0 product for us to sell. We then focused on marketing and sales to attract and convert trial customers to paying customers. Within a year, the team was 11 employees, half of whom were engineers.

What Exactly Is the Product?

As with any startup, the first major question we had to tackle was: *what would have to be added to LemonAid, as a research prototype, for it to become a salable product*? In business terms, a product is anything that satisfies a particular need or desire in a market and is sold for profit [26]. Some products are consumer-based and can be directly purchased or used by end users (described as *business-to-consumer*, or *B2C*). For example, a subscription to an online music-playing service or a mobile photo-capture application are both examples of B2C products. Other products are *business-to-business,* or *B2B*, where the buyer is another business. For example, enterprise software is a type of B2B product.

Since our help retrieval tool was designed for integration with existing websites and web applications, our product required a B2B model. We realized that it would be difficult to market our B2B product if its *only* benefit was for end users: it would have to provide some other clear value to the business integrating our tool on their site. The question was: *how can our product create value for customers who will pay us to adopt our contextual help approach?*

All of this paper's authors were co-founders of AnswerDash. The second and third authors also served full-time as CTO and CEO, respectively. They shared the responsibility of defining the *minimum viable product*, or *MVP*—the smallest set of features that would produce paying customers. As we would soon learn, our background in HCI was very relevant for identifying customer needs, translating them into feature ideas, and designing those features for engineers to build. Nearly all of our requirements emerged from conversations with customers both before and after they were paying us.

Many requirements came from the need to optimize the rate at which sales pitches and marketing efforts led to new customer signups. This included many *marketable* features that brought little end user value. For example, the ability to add video to answers was hugely popular in demos to customers, even though in practice, none of AnswerDash's current customers have yet added videos to answers. Similarly, we had to provide a mobile version of the product for customers to view AnswerDash as a "whole product," even though many businesses today still have little mobile traffic.

Other requirements emerged from needing to support existing customer support processes within our customers' businesses. We had to think carefully about how to integrate with customers' existing tools and customer support workflows, which required integrations with other services, such as Zendesk and desk.com, which many customers used for managing support requests via email.

Although LemonAid had basic content moderation and analytics during its research field deployments, AnswerDash customers immediately wanted more advanced features for managing their AnswerDash integrations. Before customers could seamlessly adopt, we had to perfect tools for styling and configuration, content moderation, answer authoring, installation and setup, and testing and analytics. Moreover, every usability issue with these features was a measurable risk to customers successfully deploying and paying for AnswerDash, and so we invested heavily in ensuring setup was fast, usable, and simple. All of our efforts, however, took time away from further focusing on end-user benefits as the original HCI research had done.

We also had to further evolve AnswerDash along many software quality dimensions. Our implementation had to be: (1) secure, so as not to risk information about our customers or their customers; (2) robust, so as to support rapidly evolving our product's features and scaling our customers; (3) reliable, so that the AnswerDash service would rarely be offline; and (4) small and fast, so as to have no discernable impact on our customers' site's loading speeds.

Overall, making the transition from our HCI innovation to a salable B2B product was more about understanding and designing a solution for which there was a *business* need and much less about solving the end user interaction problem.

Consequently, nearly 80% of the source code in the product supported customer needs, rather than end user needs. This required the founders to shift their thinking from classic HCI end-user concerns to business and adoption concerns.

Who Will Adopt the Product? Who Is the Target Buyer?

A related challenge to defining and building the product was determining which businesses would be AnswerDash's target customers, who within those businesses would purchase AnswerDash, and *why*. In our research deployments, we primarily interacted with employees who had authority over both the content of a site and its design. These team members typically included interaction designers, software engineers, content analysts, and project managers. But university-based teams may be different than industry teams.

As we explored the commercial scope of our innovation, we learned about many other types of stakeholders who were critical to adoption. The key decision-makers were directors of customer support, customer experience, customer success, marketing, and e-commerce. People in these roles, however, were not the ones using our product, but rather the target buyers paying for it. Our business-side end users included customer support staff, junior marketers, product designers, and data scientists (Figure 4). We had to keep these roles in mind, optimizing for those most often buying our product.

The stakeholders in a company varied depending on whether the website was a startup, a large corporation, or a government site. For example, at e-commerce companies, the adopters were usually directors of marketing, who were focused on increasing revenue. If they would not believe AnswerDash could impact revenue, AnswerDash would not be considered. We therefore presented revenue lifts directly from multiple successful A/B tests on customer sites. Even when we showed dramatic increases in revenue, there were sometimes other initial concerns: "we don't want another service to manage," "will it work for *us*?", and "not our pixels." Sometimes visual designers would raise concerns over visual appeal, requiring us to increase the range of style customizations available. The more stakeholders there were, the higher the chance of encountering an adoption barrier that would have to be removed through improvements to our product, marketing, and sales.

In companies that offered online applications (software-as-a-service, or "SaaS" companies), the primary adopter was usually a director of customer support. Directors of customer support viewed their job as keeping expenses down, reducing the time to resolve an email-based support request, and increasing customer satisfaction, which was usually measured by tracking "was this helpful" links in support emails and on knowledge base articles. To sell to these customers, we needed to show them how AnswerDash positively affects customer satisfaction and enables a customer support team to remain effective even as its business grows. In contrast, directors of customer success were more concerned with usability, user experience, and whether customers were successful with their website, even if it sometimes led to more support requests. Unfortunately, evaluating these metrics was often highly subjective. Directors of customer success were much less common than directors of customer support; in fact, at the time of this writing, many companies aspire to convert their reactive support organizations into proactive success organizations, and are only beginning to learn how.

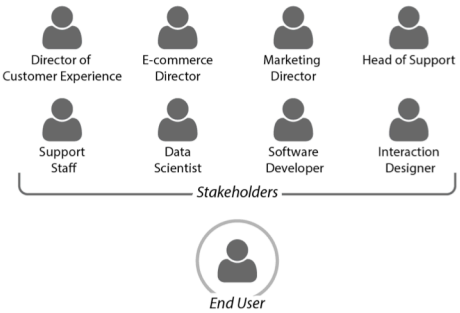

Figure 4. Typical stakeholders for AnswerDash.

In contrast to e-commerce customers, government customers were often driven by idiosyncratic, non-commercial motives. For example, some were interested in AnswerDash because it represented a way to give their constituency a voice. Others were concerned with improving the perceived quality of government information technology projects. Others still had *no* support staff, and viewed AnswerDash as a way to provide assistance with a very small impact on personnel and budget. Unfortunately, we found many government customers required support for very old web browsers, which we could not offer without unwise strain on our engineering resources.

As we worked with companies, we often encountered stakeholders with conflicting needs. For example, many customers disliked our original LemonAid modal design, especially the visual "shade" overlay, because they felt like we were "taking over their site." (Our best interpretation of this concern is that they worried that the experience felt more focused on AnswerDash than on moving their customers closer to purchasing.) However, as we found in our research studies, end users benefited from this usable modal interaction when using the product. We eventually removed the mode, causing a small amount of harm to end users' ability to find answers, but assuaging the concerns of many digital marketers and e-commerce directors, none of whom seemed to evaluate the product interactively, just visually.

Another example was in the actual content of the Q&A: end users benefit most from honest answers that transparently answer their questions, but customers often needed to write answers that avoided exposing undesirable or unhelpful truths about their products or services. We never observed any customers publishing lies, but positive spin was common in the answers our customers wrote for end users.

Other tradeoffs occurred among stakeholders *inside* AnswerDash's business customers. For example, it was com-

mon for a company to bring in leadership from both marketing and customer support when considering whether to adopt AnswerDash. When this happened, it was common that support wanted the tool to decrease costs and load, but marketing worried about losing sales opportunities from live chat, phone, and other expensive assisted service channels. Interestingly, the companies where support and marketing were most aligned were those in which leadership in customer success had equal power to marketing and support. In these companies, there was one adopter with a direct responsibility to ensure a great customer experience, regardless of what other impact it might have.

Providing Business Value through the Product

Understanding the needs of potential adopters and stakeholders was an ongoing challenge. Another challenge was developing a product that would actually provide quantifiable business value [13] to all adopters and stakeholders, beyond just providing value to the end users.

To reveal opportunities for creating business value, we had to first understand what business value meant to different stakeholders and how it could be measured. For AnswerDash, we believed our system had three potential value propositions: (1) the ability to increase sales on transactional websites, (2) the ability to decrease customer support costs, and (3) the ability to give insights into user behavior and needs, by surfacing frequent questions, along with where and when those questions were asked. All of these value propositions were plausible but untested by our HCI research on LemonAid.

After working with over 50 companies in a span of two years, it quickly became clear that end user benefit alone was not of premiere value to companies. "Improve your user (or customer) experience" was not a message that resonated with most companies, despite much hype to the contrary. On the other hand, customer support teams saw great value in saving time and doing more with less. Marketing teams responded well to messages about increased sales, and often had larger budgets. And executives were concerned with scaling their customer support team's efforts for their growing businesses without making additional investments in customer support headcount. All of these value propositions resonated, conveying business value.

In delivering this value, we had to build many new features that made our value measurable and visible. For example, we built an extensive A/B testing tool that enabled a simple comparison of the number of sales with and without our service on a site. Fortunately, over 90% of the sites on which we measured changes in sales and signups saw increases of 15-80%. In a few rare cases, sales and signups remained flat because companies published answers revealing shortcomings of their own products (*e.g.*, compatibility limitations). Such answers helped end users determine they did not want the products, but revealing this information was not always beneficial to companies, even if it was beneficial to their end users.

Other strategies we used to encourage adoption included: deploying AnswerDash on small portions of customers' sites initially, regularly communicating value through automated and manual emails, email marketing campaigns, and offering an extensive analytics dashboard that showed measurable value at login. It became clear that for trials to be successful, we needed customers to define what "value" meant to them prior to a trial. Many suggested measures they had adopted from industry best practices, but without a deep understanding of those measures' limitations. It was our responsibility to ensure measures and desired value aligned.

When confronted with AnswerDash's value, most customers adopted. But for those that did not, the prime reasons for resistance usually concerned *internal process change*. Even a small amount of additional work for employees, especially the target buyer and her team, was sometimes enough to outweigh clear evidence of increased revenues, cost savings, and time savings that customers saw had occurred. For example, one customer said that their team loved the product, the insights, the cost savings, and the benefits to their end users, but they did not want to write AnswerDash answers in addition to lengthier knowledge base articles (which were hardly ever used by comparison). It did not matter that this company was only writing one three-sentence answer per day. The *perception* of an internal process change was enough to create concerns. (It was due to such comments that AnswerDash created integrations with popular existing customer support tools, which alleviated these issues.)

Ultimately, whether a customer adopted AnswerDash only partially concerned quantifiable evidence. Despite a number of successful A/B tests, the biggest question for every customer was, "But will it work for *me*?" and "How much work will it be?" Showing customers that it would work well for them with little effort was important for success.

It is fair to say that our original research goal of providing end users with better online help has been validated across dozens of customer sites via AnswerDash, but this value to end users was never enough to drive B2B adoption. Rather, successful adoption depended on showing how target buyers would be more successful at their jobs because of AnswerDash, and in turn, produce value for their businesses in measurable bottom-line terms.

DISCUSSION

Although the extent to which any innovation becomes a widely adopted product depends on markets, customers, teams, and timing (among other factors), our experience shows that a user-centered research innovation can be the invaluable foundation of a B2B software company. In particular, our user-centered system design and evaluations that maximized realism and ecological validity helped us acquire early customers, facilitated fundraising, and provided a research prototype that directly informed the AnswerDash product roadmap. Paradoxically, however, our initial user-centered focus typical of HCI research also oc-

cluded B2B adoption issues by not revealing important insights about the real-world customer support ecosystem and stakeholder dependencies. AnswerDash has therefore spent most of its existence engaged in adoption-centered design, uncovering knowledge specific to our business and product to fuel customer acquisition and inform product priorities.

While our case study has clear limitations and there are many ways to respond to our report, here we focus on three key questions about the future of HCI technology research and adoption: (1) Is there merit in including "potential for adoption" as a criteria for evaluating research systems and their contributions? (2) How can we augment systems research evaluations with stakeholder perspectives? (3) Given the inherent lag in the successful adoption of innovations, does a research focus on adoption even make sense? Should researchers even bother?

Investigating Adoption-Centered Concerns in Research

Many researchers argue that commercialization is *not* the only, or even the best, way to have impact and that researchers should not be bothered with adoption. In fact, many researchers believe in making research contributions for the sake of advancing knowledge alone, leaving the job of commercialization to startups, open source projects, large companies, and industrial research and development (R&D) efforts.

At the same time, today's exploratory industry research labs and universities increasingly face pressure to demonstrate research impact and many encourage (or even require) tech-transfer. For example, universities are investing in setting up startup incubators, spin-offs, license agreements, and other strategies for greater research commercialization. Such endeavors inevitably require a shift to adoption-centered design, as seen in our case study. Even if commercialization is not the eventual goal, a focus on adoption might still be helpful for reflecting on design choices and research questions [26].

This raises a key question for HCI research: should some aspect of the "success" criteria for an HCI technology innovation be knowledge about its adoption barriers? In the current paradigm of HCI technology research, we are not required (or encouraged) to consider adoption as part of the system's validation. Our contributions are usually assessed on novelty, feasibility, claim-delivery, and the possibility of the system or technique generalizing to a specific user population [8]. But given the importance of adoption for potential impact, might we consider (and value) adoption and stakeholder perspectives in HCI technology research? Is there a role for research that focuses on adoption of HCI research innovations *exclusively*, assuming novelty and feasibility have already been established? Or is that not the role of HCI research? Would the HCI research community even recognize it as research?

Methods for Incorporating Adoption-Centered Concerns

If we agree that adoption and stakeholder perspectives are important, how can we adapt our evaluation methods to incorporate such perspectives? One possibility is to learn from practitioners, who may already engage in adoption-centered research. For example, interaction designers and user researchers often study target adopters, buyers, and stakeholders to inform the design of new products. These practitioners have evolved several methods to uncover stakeholder needs (*e.g.*, participatory design [23], product design walkthroughs [25], *etc.*) and methods that help align user needs with business needs (*e.g.*, [10]). Such methods orient practitioners to stakeholder needs, helping them think about the whole product ecosystem rather than optimizing a design based only on end user needs.

Adapting practitioner methods to research would require new notions of rigor, generalizability, and possibly new research communities for dissemination. These methods are also often used by practitioners on products for which there already is a market. Researchers, by contrast, often spur the creation of new markets. Also, as we discovered in our B2B sales efforts, adopters and stakeholders can actually be a moving target—uncovering such perspectives during research can be an intractable challenge. This raises critical questions about the feasibility and longevity of the knowledge these methods produce.

Adoption-Centered Concerns vs. Visionary Innovations

Even if we were to develop adoption-centered approaches in HCI technology research, some might question whether they would even apply to most HCI discoveries. After all, disruptive innovations [5] can take a long time to become successful and often initially underperform established products in mainstream markets. Why should we focus on the adoption of HCI research innovations if markets are not yet ready to accept them?

While some HCI innovations may truly be decades ahead of their time, many contributions at conferences such as CHI and UIST have a much shorter lifespan and relevance, due in part to many inventions coming from industry labs with an incentive to serve product needs or identify new product categories. It is therefore possible that a more explicit adoption-centered approach to research might increase the chances that an investor, entrepreneur, or prospective employee would see business opportunities in HCI research. Combined with other systemic changes, such as more extensive and rapid publicity of research innovations for the public and greater awareness of university intellectual property policy, an adoption-centered focus in HCI research might lead to a discipline of HCI technology transfer. There is certainly ample precedent. For example, fields such as health sciences and public health have applied a "bench-to-bedside" philosophy to the creation of the Translational Medicine discipline, which helps bridge the gap between basic and applied research and practice.

CONCLUSION

Our case study raises key questions about the link between HCI technology research and widespread adoption. Should we even be concerned about such a link? Is it sufficient to

argue that we simply should not have to be bothered with such things? Would such an attitude risk making our work less relevant? Although this case study is one of the first to provide in-depth insights into the commercialization of an HCI research innovation, we have only presented one technology, one business, one university project and one perspective. A larger collection of case studies from other researcher-entrepreneurs, particularly ones reflecting on B2C businesses, would be valuable for informing efforts to transform HCI technology research from a source of ideas to a source of commercially disseminated solutions that create widespread value. We hope that our case study will be a starting point for these important discussions.

REFERENCES

[1] Bernstein, M.S., Ackerman, M.S., Chi, E.H. & Miller, R.C. 2011. The trouble with social computing systems research. *ACM CHI-Extended Abstracts*, 389–398.

[2] Brown, B., Reeves, S. & Sherwood, S. 2011. Into the wild: challenges and opportunities for field trial methods. *ACM CHI*, 1657–1666.

[3] Chilana, P.K., Ko, A.J., Wobbrock, J.O. & Grossman, T. 2013. A multi-site field study of crowdsourced contextual help: usage and perspectives of end users and software teams. *ACM CHI*, 217–226.

[4] Chilana, P., Ko, A.J. & Wobbrock, J.O. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. *ACM CHI*, 1549–1558.

[5] Christensen, C. 2013. *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.

[6] Colyer, A.M. 2000. From research to reward: challenges in technology transfer. *ICSE*, 569–576.

[7] Foley, J. 1996. Technology transfer from university to industry. *Comm. ACM*. 39, 9, 30–31.

[8] Greenberg, S. & Buxton, B. 2008. Usability evaluation considered harmful (some of the time). *ACM CHI,* 111-120.

[9] Henderson, A. 2005. The innovation pipeline: design collaborations between research and development. *interactions*, 12, 1, 24–29.

[10] Hornbæk, K. & Frøkjær, E. 2008. Making use of business goals in usability evaluation: an experiment with novice evaluators. *ACM CHI*, 903–912.

[11] Isaacs, E.A., Tang, J.C., Foley, J., Johnson, J., Kuchinsky, A., Scholtz, J. & Bennett, J. 1996. Technology transfer: so much research, so few good products. *ACM CHI Companion,* 155–156.

[12] Jick, T.D. 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 24, 4, 602–611.

[13] Kolko, J. 2014. Running an entrepreneurial pilot to identify value. *interactions*, 21, 4, 22–23.

[14] Landay, J. 2009. I give up on CHI/UIST. *DUB for the Future blog*. Retrieved Jan 1, 2015: http://dubfuture. blogspot. com/2009/11/i-give-up-on-chiuist.html.

[15] Larsson, M., Wall, A., Norström, C. & Crnkovic, I. 2006. Technology transfer: why some succeed and some don't. *Software Tech Transfer in Soft. Engr.*, 23–28.

[16] Lee, A.S. & Baskerville, R.L. 2003. Generalizing generalizability in information systems research. *Information systems research*, 14, 3, 221–243.

[17] McGrath, J.E. 1995. Methodology matters: doing research in the behavioral and social sciences. *Human-computer interaction*. R.M. Baecker, J. Grudin, W.A.S. Buxton, & S. Greenberg (eds.) Morgan Kaufmann Publishers, 152–169.

[18] Mcgrenere, J., Bunt, A., Findlater, L. & Moffatt, K. 2010. Generalization in Human–Computer Interaction Research. *Generalization of Knowledge: Multidisciplinary Perspectives.* M. Banich and D. Caccamise, eds. Psychology Press, 277–295.

[19] Pfleeger, S.L. 1999. Understanding and improving technology transfer in software engineering. *J. of Systems and Software*, 47, 2, 111–124.

[20] Rettig, M. 1991. Nobody reads documentation. *Comm. ACM*, 34, 7, 19–24.

[21] Rogers, E.M. 2010. *Diffusion of innovations*. Simon and Schuster.

[22] Olsen, D. 2007. Evaluating user interface systems research. *Proc ACM UIST*, 251-258.

[23] Schuler, D. & Namioka, A. 1993. *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates, Inc. Mahwah, NJ, USA.

[24] Slater, S.F. & Mohr, J.J. 2006. Successful development and commercialization of technological innovation: insights based on strategy type. *J. Product Innovation Management*, 23, 1, 26–33.

[25] Szuc, Daniel. 2007. Walking through your product design with stakeholders. *UX Matters*. Retrieved Jan 1, 2015 from *http://www.uxmatters.com/mt/archives/2007/06/walking-through-your-product-design-with-stakeholders.php*.

[26] Winkler, D., Mordinyi, R. & Biffl, S. 2013. Research Prototypes versus Products: Lessons Learned from Software Development Processes in Research Projects. *Systems, Software and Services Process Improvement*. Springer, 48–59.

[27] Zelkowitz, M.V. 1996. Software engineering technology infusion within NASA. *IEEE Engineering Management,* 43, 3, 250–261.