Towards Actionable Data Science Systems: An End-user Approach

by

Ju Yeon Jung

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2022

Doctoral Committee:

Professor Emeritus John L. King
Professor Mark S. Ackerman
Professor Emeritus Geoffrey C. Bowker, University of California, Irvine
Associate Professor Steven J. Jackson, Cornell University
Associate Professor Joyojeet Pal

Ju Yeon Jung

juyjung@umich.edu

ORCID iD: 0000-0003-0076-6981

Dedication

To my family.

Acknowledgements

This dissertation, and my growth as an independent researcher, is an outcome of numerous conversations that I had with a group of great thinkers and generous people, especially my co-advisors John King and Mark Ackerman. Thanks to John for being patient with me and helping me learn how to do original research. Thanks to Mark for your insights into technology and design, which helped to shape my identity as a researcher. Thanks to Geof Bowker, Joyojeet Pal, and Steve Jackson for many invaluable comments and feedback that greatly improved my dissertation.

Thanks to everyone who made my research projects possible. Thanks to Hosun Na, Junbeom Kim, Soonyoung Bang, and Hyojin Cho for being inspiring informants and supporters. Thanks to PhD students at the University of Michigan School of Information (UMSI) for being great friends, colleagues, and fellow travelers. Thanks as well to the amazing administrative staff at (UMSI) and Rackham, who were always willing to provide support that I needed.

Thanks to staffers and patrons of Papaccino’s, Eastmoreland Golf Course, Fresh Pot, Rose City, Sweetwaters, Cafe Gabie, Yri Cafe, Starbucks Choonjung-ro, Wailea Ekahi Pavilion, Norwalk Inn, and New York Public Library for offering a place to reflect and write during my itinerant Ph.D. journey.

Last but not the least, thanks to every member of my family for being who you are, and being part of my life.

Table of Contents

List of Tables

Abstract

How can we make data science systems more actionable? This dissertation explores this question by placing end-users and their data practices, rather than data scientists and their technical work of building models and algorithms, at the center of data science systems. Inspired by phenomenological views of technical systems from CSCW, HCI, and STS, I use ethnographic and other qualitative methods to understand how participants from four studies worked with data across three settings: craft brewers producing beers, people with visual impairments engaging with image descriptions of their photos on their smartphones, and repair workers repairing broken artifacts. I analyze implications for making data science systems actionable by framing the participants as *potential end-users* of these systems.

My findings emphasize that actionability in data science systems concerns not just predictions made on mostly given datasets. Actionability in my settings arose from the ongoing work of making data relevant to artifacts and phenomena that end-users engaged with in their practices and settings. I show how this ongoing work of making data relevant was challenging. The properties of artifacts and phenomena were inherently multiple and their relevance was contingent on end-users' situations. I describe end-users' data practices as processes of "registering" (making intelligible) a contingent yet coherent set of properties to turn multiple, uncertain artifacts and phenomena into actionable versions.

My dissertation makes several contributions to emerging research on actionability and data science in CSCW, HCI, and STS literature. First, based on my findings, I theorize an approach to

data science systems that imagines actionability as driven not so much by data scientists generating predictions, or even by putting humans in the loop, but by placing end-users at the center. Second, my end-user approach to data science systems informs the technical work of data science by proposing requirements for models and algorithms to be accountable not just in their predictions but to end-users' practices and settings. Third, my dissertation integrates into data science research foundational phenomenological views from CSCW that focus on how technological systems can account for and support end-users in their domains of practice, rather than the other way around.

1 Introduction

For some years now, data science research has focused on the technical and ethical issues of predictive models and algorithms on society. As a wider range of organizations explore possibilities for integrating data science systems with their existing practices and settings, CSCW and HCI scholars have begun giving attention also to practical issues of using data science "in the wild," or in humans' real-world practices and settings (e.g., Passi and Sengers 2020; Beede et al. 2020). The practical issues of using data science systems "in the wild" can collectively be referred to as issues of *actionability*. Actionability is different from studying whether data science systems make technically valid (e.g., statistically accurate) or ethically valid (e.g., fair) predictions. Actionability concerns whether the predictions or other outputs of data science systems are practically useful and valuable for *end-users* of these systems. If studying the technical or ethical implications of data science systems concerns the effects of models or algorithms for a mostly given dataset, actionability turns attention to how the data themselves can be relevant to end-users' engagement with phenomena of interest.

In this dissertation, I seek to theorize an approach to making data science systems more actionable for end-users. How might data science systems be imagined to support the end-users of data science systems? How can the insights generated by data science systems help end-users take meaningful and effective actions regarding their phenomena of interest? To take up these questions, I start from the point of view that actionability depends on end-users being more than just participants in data science systems. The end-user approach that I seek to theorize is premised on imagining a data science system that can "adapt itself to what people can and need to do with data"

(Tukey and Wilk 1966, p.696). In short, I seek to put at the center how end-users work with their data in their practices and settings.

To develop insights into end-users' data practices that underpin my proposed approach, I bring together four studies on three settings. The studies are about the data practices of craft brewers producing beers (chapters 3 and 4), of people with visual impairments engaging with photos on their smartphones (chapter 5), and of repair workers repairing broken artifacts (chapter 6). I frame the participants in these studies as *data users* — those who themselves collect and use data to engage with artifacts and phenomena in their everyday practices and settings — as well as *potential end-users* of data science systems. "End-users" generally refer to humans who use software tools that have been designed for them. In my dissertation, I use the term in a more anticipatory and speculative manner. By framing data users as potential end-users of data science systems, I seek to explore what humans would need and want to do with data science systems based on how they work with data. By studying the data practices of these three types of potential end-users, I seek to (i) inductively understand the practices by which diverse potential end-users of data science systems *work with data* to take situated actions in their settings, and (ii) use this understanding to abductively theorize how to account for and support these data practices of end-users to make data science systems more actionable.

The inductive findings of my studies highlight several characteristics of end-users' data practices that can inform the theory and design of actionable data science systems. I describe how end-users' data practices were characterized by (i) the salience of material and conceptual constraints on artifacts and phenomena of interest (cf. statistical properties of a given dataset); (ii) end-users' views of data as a resource for taking actions in their practices and settings (cf. data scientists working with a given dataset to develop models and algorithms distant from the situations of end-users); and (iii) challenges that end-users faced in using data to access and make sense of artifacts and phenomena of interest (cf. implicit assumptions that making accurate statistical predictions of datasets leads to actionable insights into artifacts and phenomena). My findings surface how the end-users in my studies worked with data to respond to challenges of engaging with phenomena. The properties of artifacts and phenomena to which data ultimately should relate were inherently multiple. The relevance of properties to end-users' practices and settings was contingent on situations. Participants worked with their data to construct actionable versions from properties of artifacts and phenomena in a situation. Working with data involved processes of situating, selecting, and putting together data about relevant properties of artifacts and phenomena in ways that were materially and conceptually coherent.

My dissertation makes several contributions to emerging research on actionability and data science systems in CSCW, HCI, and STS literature. First, I propose an initial set of three requirements for actionable data science systems. The actionability of a data science system depends on: (i) the *relevance* of the data for end-users (do data capture the relevant properties of artifacts and phenomena of interest, and the relevant relations among these properties?); (ii) the *flexibility* of data use for end-users (does the system enable end-users to flexibly use data to interact with artifacts and phenomena from different points of view and across different situations over time?); and (iii) the *accountability* of systems in the context of end-users' practices and settings (can the data science system properly account for how end-users experience and engage with real-world artifacts and phenomena as part of their practices and settings?). Since these requirements were developed by examining less-standardized and resource-constrained settings where end-users collect and work with data themselves, the proposed requirements for actionability may open up opportunities to make data science systems more accessible across a broader range of settings.

Second, my proposed end-user approach to data science systems may provide useful insights for the technical work of data science by helping develop statistical models and algorithms that better align with what end-users want to do with their data in their practices and settings. Similarly, it could help guide large institutions and corporations to improve the actionability of their data science projects by tightly integrating end-users (e.g., domain experts) with the work of their technical data scientists.

Third, my dissertation contributes theoretical insights for the CSCW, HCI and STS literature on data science regarding the relationship between data, humans, and artifacts. Data science research has often viewed data as abstract, objective representations of phenomena that can be analyzed statistically by data scientists, or as the result of human work that draws on a range of different values, intentions, and practices. The end-user approach that I develop from the four studies draws from a long line of research in CSCW, HCI and STS emphasizing a phenomenological view of humans' engagement with technology.

In this phenomenological view, understanding the effects of technology depends on accounting for how actions are situated in humans' (e.g., end-users') real-world practices and settings (Agre 1997; Smith 2019; Suchman 2007). For data science research, a primary implication is that a data science system should account for how humans use data to "register," or make intelligible, artifacts and phenomena (Smith 1998). Registration means not just statistically mapping input data to output data, or even documenting humans' interpretations of the data. Registering means mapping data to artifacts and things and their relevance to phenomena. Emphasizing registration enables understanding the data in data science systems as contingently *presented* (rather than "re"-presented) to end-users in a situation.

In the remaining part of this introductory chapter, I first give background on how actionability has been discussed in data science research. I then illustrate the trajectory of how the four studies included in this dissertation unfolded and gradually informed one another over the course of my doctoral studies, eventually leading to a proposal for an end-user approach for making data science systems more actionable. I end this chapter with an overview of the rest of the dissertation.

1.1 Actionability in data science research

Issues of actionability have recently begun to surface in data science research, as many organizations report challenges of developing and deploying data science systems as intended. Issues related to actionability include the messiness of real-world datasets that limits the predictive accuracy of models and algorithms (e.g., Sambasivan et al. 2021; Beede et al. 2020), the opacity of algorithms that harms end-users' trust in the outcomes generated by data science systems (e.g., Passi and Jackson 2018; Burrell 2016), the lack of relevance of predictions to domain-specific knowledge and experience (e.g., Athey 2017), and the difficulty of integrating data science systems with existing practices and systems (e.g., Davenport and Ronanki 2018).

While actionability has only recently emerged as a salient topic in CSCW, HCI and STS literature on data science, it has been central to the very idea of data science from its inception. Technical data science research, for example, has commonly understood data science as the systematic analysis of large datasets to generate actionable insights into complex problems (e.g., Dhar 2013; Barnaghi et al. 2013; Krause et al. 2016; Cao 2017; Taieb 2018; Ustun et al. 2019; Verma et al. 2020; Karimi et al. 2021; Sarker 2021).

While the Oxford English Dictionary defines actionability as the ability "to be acted upon or put into practice; useful, practical" — indicating the importance of *relevance to practice* —

both the data science literature and common understandings of data science imply a narrower view of actionability. Action in real-world practices and settings is subsumed by a concern with the models or algorithms used in data science (e.g., improving their performance or making them more understandable). Data are usually understood as they relate to data scientists' technical practices of making these models and algorithms work. Actionability is discussed in terms of, for example, the explainability and interpretability of models or algorithms. The implicit assumption is that making models and algorithms explainable and interpretable to humans can make data science systems actionable in practice.

Existing research discusses explainability in terms of mechanisms such as *post-hoc* explanations of an opaque model, integrating explanations, explaining the general logic of a new model, or explaining a particular outcome (i.e., prediction) of a model (e.g., Guidotti et al. 2018; Wieringa 2020; Rudin 2019; Wang et al. 2019c). These existing views of explainability are mostly *post hoc*. These views of explainability also do not necessarily improve even post-hoc actionability in a data science system. Krause et al. (2016), for example, show that, although visualizing partial dependencies between features helped users explore the models that they had developed, many highly predictive features were either inherently not changeable (e.g., age) or not practical to change (e.g., while the risk of diabetes was highly correlated with the number of lab tests, recommending patients to get fewer tests to lower the risk would not make sense).

Other researchers have pointed out that an explanation of how a system reached a particular outcome is not necessarily sufficient for making explanations actionable. They advocate for "actionable" explainability that focuses on designing algorithms to generate explanations that can help users achieve desired outcomes (e.g., Wachter et al. 2017; Ustun et al. 2019; Verma et al. 2020). The argument is that a data science or AI system should be able to generate "counterfactual explanations" that suggest actions to achieve a more favorable outcome (e.g., for a credit loan algorithm, the user should be able to query, "How can I get my next application approved?") (Wachter et al. 2017; Singh et al. 2021). To improve actionability, this research explores what could be accounted for in algorithms, such as the actionability of input features (e.g., a user cannot change their age or race) (Ustun et al. 2019), real-world causal relationships among features (Karimi et al. 2021), or different types of explanations (e.g., suggesting a generic action or specific actions) (Singh et al. 2021).

While explainability and interpretability are often used interchangeably in data science, several scholars stress the need to differentiate the two concepts. Gilpin et al. (2018) argue that interpretability concerns making the inner workings of ML models understandable and thus trustworthy to humans (also Coussement and Benoit 2021); an explainable model is not only interpretable but also able to defend its outcome. In this view, explainability assumes interpretability, but not *vice versa*. Doshi-Velez and Kim (2017) argue that ML models need to be interpretable only when the problem is under-studied or the outcomes of models have non-trivial negative consequences. Rudin and Radin (2019) (also Rudin 2019) make a more fundamental criticism of the existing approaches to interpretability that take the opacity of ML models as given. They argue that simpler, inherently interpretable models (e.g., relatively simple decision trees versus complex deep learning nets) may be more effective and ethical than opaque models that can only be made understandable *post hoc.*

1.1.1 Actionability as model-driven

Existing literature on actionable data science makes several underlying assumptions that do not explicitly focus on actionability from the end-user approach that I seek to explore in this dissertation.

First, focusing on the explainability or interpretability of models risks reifying data science as driven by models abstracted from practice rather than humans' actions situated in the real world. As Shneiderman (2020) comments, conventional data science and AI discussions conceptualize that "humans are "in-the-loop" around AI," rather than "AI is "in-the-loop" around humans" (p.112). Extending this critique, my dissertation will propose an approach to actionable data science in which end-users' data practices are at the center of data science systems.

Second, algorithms for actionable explanations tend to build on a largely pre-defined set of features or relationships among features. The assumption is that the world that the data capture is largely stable, in the sense that the statistical distribution in the initial model is assumed to hold even when more data come in (Marcus 2018; Athey 2017). Algorithms are "consistent" in this statistical sense, and also "scalable", leading to more "consistent decisions" (Raghavan 2021, p.2). The nature of artifacts and phenomena is implicitly taken as given before humans actually engage with them when using a data science system. I draw on the concept of registration — a process of humans making an artifact in the material world intelligible, or able to "speak and think about it, act and conduct their projects" (Smith 2019, p.66-7) — to highlight how (i) our ontological and epistemic access to artifacts and phenomena in the world is not given but a *result* of registration activities, and (ii) real-world artifacts and phenomena themselves are inherently multiple, uncertain, and contingent.

Lastly, in their focus on making algorithms' outcomes (i.e., predictions) actionable, existing approaches to actionable data science do not focus so much on the role of the end-users of data science systems. When predictive models can extract statistical relationships from datasets accurately (e.g., A will lead to B or A is positively correlated with B) (Argyris 1996), data science techniques can often be "successfully applied by data scientists with little knowledge of the problem domain" (Athey 2017, p.483). Here, actionability is tied to improving the predictive accuracy of algorithms. Yet end-users' role in this research relates largely to the work of collecting datasets or arm's-length informing of data scientists. End-users take up a peripheral position in data science research and practice, with data scientists at the center (Gil et al. 2019; Amershi et al. 2019b; Sambavisan and Veeraraghavan 2022).

A less-examined question in data science research and practice remains: what makes the insights that emerge from data science systems actionable for their end-users? Recent data science research has focused on settings where end-users need not be the producers or collectors of data used in building data science systems, or where the artifacts or phenomena of interest are highly standardized or well-defined. The everyday practices and settings of the end-users of data science systems, from which the data for data science ultimately come and in which data and models are ultimately used, are usually left mostly invisible (Sambasivan and Veeraraghavan 2022; Kitchin 2014; Smith 2019). My dissertation frames humans who collect, manage, and use data (i.e., data users) in their practices and settings as potential end-users of data science systems. I seek to show how this end-user approach may open up novel opportunities to make data science systems more actionable.

1.2 The trajectory of a multi-case study

The four studies in this dissertation cover diverse contexts and settings (e.g., craftwork, brewing, data science, visual impairments, photography, and image descriptions, computer vision, repair, artifacts). What guided me from one study to another were happenstances and hunches. Over time, the themes of end-users, data practices, and data science systems, which together developed into my proposal for an end-user approach to data science, emerged through an iterative cross-analysis of different subsets of the four studies. I have arranged the studies retrospectively to construct a narrative account of an end-user approach to making data science systems more actionable.

Out of the four studies, the ethnographic study on repairedness (chapter 6) was conducted first, in the summer of 2019. Up to that point, my main research interest concerned how artifacts, especially photos, could be organized to support ongoing meaning-making. I had previously explored how humans (e.g., bereaved parents and people with visual impairments) tried to make meaning out of and manage their personal photo collections. The repair study was initially conceived as an extension of this prior research on how humans make sense of, use, and organize a variety of artifacts (e.g., many different photos in a photo archive and many different repair artifacts encountered in a repair workshop). I chose Sewoon Sangga, a historic analog electronics repair community in Seoul, as the fieldsite based on my prior familiarity with the variety of artifacts the workers dealt with at the site. The site was characterized by cluttered physical spaces in the community (imagine a repair worker fixing a machine surrounded by piles and clusters of parts, tools, and partially disassembled items) and contingency-ridden repair work processes (not being able to predict what items will be received when and in what conditions, requiring what kinds of resources). During the fieldwork, I was struck by how the artifacts under repair that were mostly antiquated or custom-built were hardly fully or objectively repaired according to a predefined set of requirements. The repairedness of artifacts was negotiated at the level of their *properties* (e.g., a radio with a sufficient volume level or an electric pipe organ whose boot time is reasonably short but without any memory problem) and artifacts became *contingently stabilized* into a version that was considered *sufficiently* working for the owner of the artifact at a given time. Over time, whether or not there were explicit databases involved, I found it useful to understand the properties of the artifacts in this setting as properties of data about the repair workers' practices and settings.

In 2020, I happened to meet the founder of a craft brewery in Korea, who was interested in "systemizing" how the brewers were collecting and managing their brewing data. After a series of casual conversations, we decided on an exploratory research project at the brewery (chapter 3). My goal was to understand the brewers' data practices and possibly derive some recommendations. The brewers' data practices seemed to be an interesting opportunity to explore how the brewers made sense of and gained a sense of control over the properties of their artifacts in the context of the phenomenon of interest (e.g., brews, equipment, ingredients in the brewing process). While I was not initially looking for any direct connection between the repair study and this proposed study at the brewery, the uncertainty of their central artifact (it's hard to fully define what a brew is supposed to be, directly see what's going on with it during the brewing process, or predict how it will turn out) and thus the inherent multiplicity of artifacts (e.g., Law 2002; Mol 2002) seemed to be an underlying theme that continued in this new setting.

From the beginning of the brewing project, I had interest in studying the potential relationship between the brewers' situated data practices and technical data science activities. Brewing was data-intensive (involving collecting and reviewing data on hundreds of variables throughout the grain-to-glass brewing process). The brewers were collecting and managing their data mostly manually using paper brewsheets within the physical brewery. The fieldwork led me to see that (i) the brewers needed to situate their data in a particular brew, brewing process, or brewery setting to make sense of and meaningfully reason with the data; (ii) they wanted to use data to make decisions about what actions to take throughout the brewing process rather than getting a prediction on the outcome of the process; and (iii) they prioritized the ability to flexibly account for and address unpredictable contingencies and anomalies in data over the ability to achieve a target value precisely. The study stressed that developing data science systems for craft brewers should account for how the brewers used data distinctly from technical data scientists in terms of their view of data (situated vs. abstract), purposes for engaging with data (guiding processes over predicting outcomes), and overall goals of using data (flexible control vs. precision).

The fieldwork continued and, by early 2021, my key research interest had become how domain experts' data practices could inform theories of the activities of data science systems. To further explore the relationship between data practices and data science activities, the brewery allowed a pilot machine-learning (ML) project at the site with a data scientist (chapter 4). The aim of the project was to develop a predictive model for yield efficiency (the amount of beer produced relative to the grain used) using the dataset collected by the brewers. The dataset was incomplete and unreliable, limiting the effectiveness of data science activities as well as the actionability of predictions. While the problems with messy real-world datasets were well-established in existing data science research in terms of data wrangling, I questioned how and why the dataset became incomplete and unreliable in the first place, directing attention to, again, the brewers' data practices.

Focusing on the brewers' data practices helped me identify several situations that led to unreliable, incomplete data, and thus to sources of misalignment between domain experts' data practices and data science activities, specifically of domain experts. The study argued that, for data science systems to be actionable in a sustainable way (e.g., integrated into everyday work practices and settings), the systems need to go beyond focus on improving the predictive accuracy of algorithms to (i) support domain experts' different uses of data, (ii) enable domain experts to see the value of data science systems and the importance of collecting complete, reliable data, and (iii) support flexible management of datasets as they are collected by domain experts.

The two studies at the brewery led me to develop a view of data science driven by domain experts, rather than by data scientists, where domain experts could be viewed as potential end-users of data science systems. I began to imagine and explore data science systems as technologies that are (i) developed from datasets shaped by domain experts' data practices in their practices and settings, and (ii) to be used by domain experts to engage with phenomena of interest through their data.

During this period, I started to explore how my findings from the repair and brewery settings could be integrated. A common finding seemed to be that both the repair workers and the brewers used data to access, make sense of, and act on the *artifacts* or *phenomena* of interest at the level of their properties: What is this thing (e.g., a problematic electric pipe organ or a brew) or what is it supposed to look like (i.e., what aspects, functions, or variables to examine and which values or statuses do they see as proper)? How can I make an effective intervention to successfully get my job done (e.g., what to do to make the organ work again or make beers consistent in quality)? The relationships between humans as situated data users, data, and properties of artifacts and phenomena emerged as a possible umbrella theme of my research.

As this theme was being developed, I returned to photo use and archiving, on which I had done research early in my doctoral studies. This time, I was especially interested in how people with visual impairments access, make sense of, and engage with their photos. Working with a visually-impaired artist, who had been a long-time informant, I took on an interview study on the use of personal photos by people with visual impairments (chapter 5). What this setting had in common with the previous settings, I began to realize, was the inherent uncertainty of the artifact or phenomenon of interest. People with visual impairments had no or limited direct access to the photos as visual artifacts, and they needed to draw on image descriptions or other data to access, make sense of, and do things with their photos. As my interviews continued, participants' experiences of and expectations for image descriptions (both computer- and human-generated) emerged as a key theme. They needed to construct a mental image of the photo that they wanted to engage with drawing on available data. The process of constructing a mental image differed across photo activities and goals. I concluded that computer-generated image descriptions should be able to account for these differences in photo activities and goals to help people with visual impairments engage with their photos as they intended.

This study also indicated to me the need to bring to the center the potential end-users of a data science system and their engagement with data in their phenomena of interest to make these systems actionable. A question lingered regarding the relevance of repair workers to my working concept of actionable data science driven by end-users' data practices. Could repair workers be seen as potential end-users of data science or data-driven systems? My conclusion was that, even if they cannot be seen as end-users of data science systems *per se*, they could be seen as data users — as users who work with a wide variety of data in their practices and settings (e.g., sensory data, measurements displayed on testers, parts numbers and specifications, inventory data on parts and tools, etc.). The study was critical to the end-user approach that I was developing since it offered me an account of how humans use data to turn an inherently multiple and uncertain artifact into an actionable version. The study contributes insight into the role of artifacts and phenomena in the practices of *data users*, which served as the basis of the end-user approach to actionable data science that I developed.

While dealing with diverse contexts and settings, the four studies ended up contributing to the theme of this dissertation of making data systems more actionable by offering insights into end-users' data practices. The research process that I took may be seen as a multi-case theory-building approach (Eisenhardt 2021) that develops a set of theoretical constructs and their connections grounded in the iterative analysis of empirical findings of different cases to probe a phenomenon "for which there is little or conflicting prior theory and/or empirical evidence" (p. 148). Choosing different cases can be seen as a type of theoretical sampling to improve both the "creativity and reliability" of the argument (p. 152). While some of the studies in this dissertation were not initially considered to be directly relevant to data science as we know it, the iterative process of analyzing and combining the theoretical themes of the different cases enabled me to develop a novel approach to data science that might not have been possible had the research involved more restrictively defined or narrowly selected cases.

1.3 Overview of the chapters

The rest of the dissertation is structured as follows. In chapter 2 (**Data and actionability in data science research**), I draw on existing literature in CSCW, HCI, STS, and the philosophy of information to analyze assumptions about data and actionability in existing data science research. The analysis focuses on (i) the relationship between data and phenomena, (ii) how data become actionable; and (iii) proposals to make the data in data science systems actionable. I point out that what remains largely invisible in the existing assumptions of data is how end-users of data science systems collect, manage, and work with data in their practices and settings. I sketch out what I propose as a phenomenologically informed view of data and data science systems that accounts for relationships between data, data users, and phenomena as they are situated in practices and settings. This proposed phenomenological view of data and data science systems serves as a basis for analyzing and integrating the findings that I lay out in the next four chapters on data practices across different settings.

In chapter 3 (**How domain experts work with data**), I examine the situatedness of data and data use in an ethnographic study at a craft brewery. The chapter explores a domain expert-driven understanding of data science through the lens of craftwork, analyzing how a team of craft brewers used data to produce high-quality beers as materials and circumstances in their brewery continually shifted. The vignettes included in the chapter illustrate how the brewers worked with data by situating changes in the values, target ranges, and relationships of variables within their practices and settings. By working with data, the brewers were able to gain a sense of control over complex and unpredictable materials and brewing processes. The study contributes theoretical insight into how domain experts' (or end-users') use of data is distinct from the use of data by technical data scientists. It also contributes practical implications for CSCW and HCI research on designing tools that support end-users' engagement with data science activities in their domain-specific work practices and settings.

In chapter 4 (**Understanding domain experts' data practices for sustainable data science**), I draw a connection between data users' data practices and the ability to sustain data science activities over time. The chapter illustrates the course of a pilot machine learning project at the same craft brewery. The focus is on exploring the relationship between the brewers' data practices and the data science activities that depended on that dataset. The chapter describes three types of situations where brewers' data practices led to unreliable, incomplete data, and how such data practices limited the effectiveness of data science activities. Drawing on analysis of the sources of misalignment between the brewers' data practices and data science activities (the brewers' logical uses of data vs. the data scientist's statistical use of data; the unclear value of reliable, complete data to the brewers; and difficulty to manage data as they are collected), the chapter discusses design implications for sustainability in data science activities. Inspired by research on end-user software development that views sustainability as driven by domain experts as *owners of problems*, the chapter proposes data science research driven by domain experts as *owners and users of data*.

In chapter 5 (**Expectations of people with visual impairments for image descriptions**), I present a case where data users have little or no access to the properties of artifacts of interest - people with visual impairments (PVI) accessing photos through image descriptions. Thus far, research has studied what PVI expect in these descriptions mostly regarding functional purposes (e.g., identifying an object) and when engaging with online, publicly available images. Drawing on interviews with 30 PVI, the chapter examines their expectations for image descriptions when viewing, taking, searching, and reminiscing with personal photos on their own devices. It shows that they needed to use data to construct mental images of the content of a photo to make it actionable. Their expectations of what data should be included in descriptions and how the descriptions could be delivered varied across different types of photo activities. How they wanted to engage with photos through descriptions went well beyond identifying objects in photos. Based on the findings, the chapter proposes design opportunities for generating and providing image descriptions for personal photo use by PVI. The study stresses how these participants saw data as clues to be arranged or put together in a particular way to construct a mental image through which they could experience the photo.

In chapter 6 (**How artifacts under repair become contingently stabilized**), I bring to the center the role of artifacts in humans' data use. The chapter focuses on the following question: When does an artifact under repair become at least contingently stabilized as working again?

Drawing on a series of vignettes from ethnographic fieldwork at an analog electronics repair community in Seoul, South Korea, the chapter examined "repairedness" — the contingently stable, working version of an artifact under repair that is negotiated out of multiple possible versions to bring about the temporary conclusion of repair work. The repair workers and owners negotiated versions of the artifacts by strategically identifying and prioritizing particular properties of the artifacts as well as determining the sufficiency of these properties in a given repair situation. These properties were scoped, re-prioritized, and temporarily substituted during the negotiation process, shifting what constituted a working version of the artifact under repair. While the study did not explicitly mention "data," repair workers' engagement with properties of artifacts could be usefully understood as how they selected and put together data about those properties in their given situation (e.g., based on the numbers shown on a voltage meter's LED screen; the level of volume of the radio, the boot time of an electric pipe organ). The process of negotiating working versions of artifacts spotlighted challenges for using data to turn contingently relevant properties of artifacts into actionable versions.

In chapter 7 (**Toward actionable data science systems: an end-user approach**), I summarize and analyze the four studies (chapters 3-6), focusing on commonalities across their settings as well as in the findings as they relate to data science research. I then propose an end-user approach to data science, including an initial set of requirements for actionable data science systems of *relevance*, *flexibility*, and *accountability*. In chapter 8 (**Conclusion**), I sketch out future research directions to explore how the end-user approach that I propose may be drawn on by data science researchers and practitioners. I suggest four types of end-user data practices to support (flexible and efficient data annotations, exploring logical and material relationships in data, generating intermediate predictions, and developing data science strategies) to make data science systems more actionable for end-users. I conclude with a brief discussion of limitations.

2 Data and Actionability in Data Science Research

Data power and shape data science systems (Amershi et al. 2019b). Yet data science researchers are only recently beginning to explicitly acknowledge the critical role of data (Gebru et al. 2021; Sambasivan et al. 2021), with some proposing a need for a turn to "data as the object of study" (Ribes 2019, p.515; Alaimo and Kallinikos 2021). Studies of data science have tended to use the term "data" casually or define data only implicitly through examples (Borgman 2017, Meadows 2001, p.3). Kitchin (2014b) notes: "it is remarkable how little conceptual attention has been paid to data in and of themselves. In contrast, there are thousands of articles and books devoted to the philosophy of information and knowledge. Just as we tend to focus on buildings and neighbourhoods when considering cities, rather than the bricks and mortar used to build them, so it is the case with data" (p.1).

To study the actionability of data science systems, I suggest beginning by exploring possible concepts of *data* and their relation to data science. The Oxford English Dictionary defines "datum" as: (i) "an item of (chiefly numerical) information, especially one obtained by scientific work, a number of which are typically collected together for reference, analysis, or calculation"; (ii) "something given or granted; something known or assumed as a fact, and made the basis of reasoning; an assumption or premise from which inferences are drawn"; (iii) (philosophy) "anything immediately apprehended by or presented to the mind or senses." In the academic literature, the diverse dictionary definitions show up in tensions between the view implicit in much data science research of data as given ((ii)), versus social scientists' emphasis on how data are constructed ((i) and (iii)).

Drawing on the tensions in these definitions, as well as literature in CSCW, HCI, STS, and the philosophy of information, I next explore different assumptions of data. My analysis will focus on (i) the relationship between data and phenomena, (ii) how data become available and actionable, and (iii) existing proposals for making data in data science systems actionable. To extend existing assumptions of data and their actionability, I propose a phenomenologically informed view of data. Using this view, I conceptualize data, data users, and artifacts and phenomena for the purpose of this dissertation. These concepts will inform my analysis of the four studies presented in chapters 3-6.

2.1 The relationship between data and phenomena

Much existing literature in CSCW and STS understands data as being in relation to concrete phenomena in the world (e.g., Shah 2013, Kitchin 2014b; Leonelli 2015). Data are described as, for example, "entities" used as "evidence of phenomena" (Borgman 2017, p.29; Bogen and Woodward 1988), or as a "surrogate, adequate or inadequate, for some phenomena" (Drucker 2018, p.249).

There exist differing views on the extent to which data, once collected, can be separated from phenomena. Drawing on studies in HCI and STS, Offenhuber (2020) summarizes existing views of data according to two axes: (i) an epistemological-ontological continuum between "data as a product of the human mind versus data as observer-independent patterns in the world" (p.28) and (ii) a representational-relational continuum that compares assumptions of data as unambiguous references to a phenomenon (i.e., data correspond to fixed features of a phenomenon), to embedding a set of complex and dynamic relations between humans and the phenomenon (i.e., how data relate to a phenomenon shifts through various types of human intervention within a sociotechnical system).

Views of data as ontologically unambiguous, objective representations of a phenomenon are common in the technical data science literature. In this literature, data are often assumed to be a "raw material" for data science (Kitchin 2014b, p.1). Data are "readymade" representations (Preissle and Han 2012, p. 595) that, once abstracted from a phenomenon, can be treated, explored, and made sense of in their own right as objective components of the phenomenon, separate from, or in the absence of, the phenomenon (e.g., exploring patterns within a dataset on a computer) (Clough et al. 2015; Offenhuber 2019). Phenomena are assumed to be complex yet stable. The relationship between data and phenomena is therefore reliable, and the relevance of data to phenomena is mostly given (Marcus 2018; Drucker 2018).

While representational views of data can be efficient and productive in that they make data analysis generalizable and scalable beyond the original phenomenon, relational approaches emphasize the situatedness of data collection and use (Leonelli 2015). STS and CSCW literature emphasize relational views of data. Data are not ready-made representations of stable phenomena, but the result of articulating and negotiating relations surrounding phenomena in particular contexts (Bowker and Star 1999; Kitchin 2014b; Porter 1995; Pine and Liboiron 2015). Data cannot be considered separate from a complex and dynamic network of sociotechnical factors, making data and phenomena mutually constitutive (Gitelman and Jackson 2013). The relation between data and phenomena is neither objectively given nor fixed. Dealing with data requires awareness of value-laden human decisions. Engaging with phenomena through data is an inherently social activity. In the following two subsections, I discuss these different assumptions about the relationship between data and phenomena further.

2.1.1 The separation of data from phenomena for statistical analysis

Computers were developed as artifacts to process and deal with the world as representations. How to represent things, activities, and environments in a way that computers could recognize and process has been regarded as a critical task in developing computational tools (Agre 1997). Turning phenomena into representations is a dynamic process that entails focusing on particular aspects of phenomena while hiding others at various levels of abstraction whose relevance depends on the context in which the data are used (Nadeem 2022; Offenhuber 2020). Data science literature often assumes a view of data as representations that ‘correspond to’ or ‘designate’ ‘features’ of a phenomenon in the real world (Offenhuber 2019; Royal Society 2012; Leonelli 2015). The relationship between data and phenomena is thus rather fixed and context-independent (e.g., a percentage displayed on a battery level indicator objectively corresponds to the battery’s current state of charge, regardless of the goals and intentions of a human user). As data objectively correspond to the features of a phenomenon, a dataset itself is seen to be enough to gain access to and engage with the complexities of the phenomenon (O’Sullivan 2018). Complex statistical or computational analysis can be performed on the given dataset, where the insights generated could be safely assumed to be relevant to the phenomenon.

As in the example of the battery level indicator, assuming a correspondence between data as they are represented and the features of a phenomenon as they appear might not be problematic in many cases. With proposals to make data science a “*universal science*” (Ribes 2019, p.515, emphasis original), however, assumptions of data as objective representations have led to a view that almost everything - from social trends to individual behaviors and identities - can be represented and managed as data, or amenable to being “datafied” (Cheney-Lippold 2017). Once captured and standardized, data could be “unmoored” not only from the real-world contexts surrounding the creation and use of data (Kitchin 2014b, p.22). Once unmoored from their context, data can be manipulated and analyzed computationally and statistically (Sloane et al. 2020; boyd and Crawford 2012). A view of data as decontextualized representations of phenomena makes it possible for the object of investigation to shift from phenomena in the real world to given datasets on a computer, separate from and independent of the phenomena. What remains largely invisible in this view are how data are related to phenomena and how humans use data to engage with a phenomenon as it unfolds in the real world (Offenhuber 2020; Smith 2019).

2.1.2 Achieving the fit between data and phenomena

As data science systems increasingly permeate high-stakes decisions (from healthcare to employment and law enforcement), scholars of CSCW and STS have begun to explore the ethical implications of viewing data as decontextualized abstractions (e.g., Rudin 2019; Sambasivan et al. 2021; Gebru et al. 2021; Veale et al. 2018). Their research stresses that the fit between data and phenomena is not given but to be worked toward and achieved by human users in a given context (Taylor et al. 2015; Feinberg 2017; Gitelman 2013; Aragon et al. 2022). This research problematizes statistical measures of the fit between data and models, such as predictive accuracy. The argument is that data science should aim for improving the fit between data and real-world phenomena in terms of "fidelity" (how closely data represent phenomena) and "validity" (how well data explain aspects of the phenomena that data present) (Sambasivan et al. 2021, p.10). Existing work makes this argument both in pragmatic and sociotechnical terms, and using theories of fairness.

First, a view of data as contextualized has arisen pragmatically from the observation that improving the fit between data and phenomena depends on human involvement to document, keep track of, and communicate the context in which the data is situated. Advocacy of the 'bigness' and 'unmooring' of data may lead to "documentation debt" (Bender et al. 2021; Gebru et al. 2021).

Dealing with large datasets requires careful planning to make visible the relationships between data and phenomena.

Second, data are embedded in sociotechnical systems and created and managed through human work. Data are interpreted not only by *post-hoc* "wrangling" or exploratory data visualization using statistical techniques. Data are shaped by human actions that extend to the initial data collection and to the use of data in real-world practices and settings. These human actions shape the relationship between data and what aspects of phenomena can be known and how (Pine and Liboiron 2015; Passi and Jackson 2017; Kitchin 2014a; Star and Bowker 1999). Data are ontologically and epistemologically reflective of different values, intentions, and goals of stakeholders (e.g., Passi and Sengers 2021; Passi and Jackson 2018). In contrast to viewing data as decontextualized, this research from CSCW and STS focuses on making visible the actions and social processes of constructing data and enacting data science practices.

Third, this research also emphasizes how data shape not only the statistical performance of models and algorithms but also the fairness and robustness of statistical predictions (Gebru et al. 2021). Research in HCI and CSCW has started to examine data science implementation problems across high-stakes settings such as healthcare (Sambasivan et al. 2021; Beede et al. 2020), criminal justice (Chancellor et al. 2019), and finance (Passi and Barocas 2019). Mismatches between data and phenomena become a serious social problem when applying data science techniques to such high-stakes issues. In their study of data practices in high-stakes AI, Sambasivan et al. (2021) reported that the effectiveness of data science and AI practices is severely curtailed by poor-quality data, leading to what they refer to as "data cascades" or "compounding events causing negative, downstream effects" (p.5) on the performance of models and algorithms.

2.2 How data become available and actionable

Data also generate conceptual tensions regarding how they become available and actionable to humans (and computers). The etymology of the term data is "what is given" (Leonelli 2015, p.811) or "that which is given prior to argument" (Rosenberg 2013, p.36). Data, then, are already somewhere out there for us to use, prior to or independent of humans' direct engagement with a phenomenon. Drawing on historical analysis, Rosenberg (2013) concludes that this givenness embedded in the term data has allowed it to perform a uniquely "rhetorical" role (p.18), without necessarily being tied to any ontological or epistemological claims – that data are taken to be data, regardless of their meanings, quality, or validity. Data are "pre-analytical" (Kitchin 2014a), "uninterpreted," or "proto-epistemic" (Floridi 2011, p.85). In this view, data are seen as raw material to be processed through iterative rounds of interpretation and analysis into valuable and meaningful information or knowledge about the phenomena that data are assumed to represent. Here, actionability is more about the process of converting data into something else (information or knowledge) than about bringing data to bear on phenomena directly.

In CSCW and STS, actionability can be connected to the frequent emphasis on action as situated. Data are purposefully created and collected through situated action with the world (e.g., through observations, experiments, computational processing). Data do not exist prior to humans' interactions with phenomena in the world but are the outcome of such interactions (Leonelli 2016, p.192, Star and Bowker 1999). Data production and use are always situated in particular practices and settings (Taylor et al. 2015; Lave 1984), and what counts as data is therefore always "relational" (Leonelli 2015, *passim*). Based on their ethnographic study of the data-related activities of an urban street community, for example, Taylor et al. (2015, p.2864, emphasis original) propose the term "data-in-place" to "think of [data] in terms of a *social geography* in which data, people, and things intermingle to continuously enact place." Borgman et al. (2012) in this context examine a scientific infrastructure project involving environmental scientists, computer scientists, and engineers, and find that their collaboration depended on negotiating different perspectives on what constitutes data. These studies stress the situatedness over the givenness of data. Data exist and serve as a resource for producing knowledge or part of a knowledge production system (Leonelli 2016; Borgman 2017).

What does this literature imply about how data are made actionable? First, if humans' actions are situated, then actionability depends on selecting data for particular situations. Etymologically, the emphasis turns to how the term data is a plural form of datum. It follows that the data that relate to a phenomenon are assumed to be considered collectively and "clustered together" according to given rules and systems (Floridi 2011, p.84). Any collection of data, as Kitchin (2014b, p.2) argues, is "inherently partial" and "always a selection of the total sum of all possible data available." A set of data is intentionally selected when they are perceived as potentially relevant to a situation within some phenomenon of interest by a particular human user (Kitchin 2014b; Borgman et al. 2012; Leonelli 2016). This emphasis on the selection of data has consequences for actionability in that selecting different subsets of data leads to different models and different real-world actions (Gitelman and Jackson 2013).

In more technical approaches (e.g., in machine learning), selecting data concerns the tasks of feature selection and feature extraction. Feature selection refers to the process of selecting relevant features of data (or removing irrelevant features) to be considered for modeling and analysis; feature extraction usually involves the transformation of given data to a set of features in a way that preserves their relevance to a particular problem (Khalid et al. 2014). In machine learning, the purpose of these tasks is to reduce the dimensionality of the data to improve statistical properties of the model (e.g., accuracy, bias, variance) as well as understandability of results for human users (Blum and Langley 1997). ‘Relevance’ in feature selection and extraction in machine learning is used in a statistical sense. Feature selection here is considered an optimization problem to construct a subset of the “least number of dimensions that most contribute to learning accuracy” (Khaild et al. 2014, p.372; Cai et al. 2018). This statistical view of feature selection contrasts to approaches that view the relevance of and selecting of data as a matter of negotiating and making decisions about a given dataset by humans to qualitatively align the features to their own goals and values. In the following two subsections, I explore what these differing approaches have to say about how humans make data and data science systems actionable.

2.2.1 Making data actionable through models and algorithms

In technical data science research, data are explored and analyzed mainly through *data scientists*’ use of statistical models and algorithms. The ‘insights’ come from novel or subtle patterns within a dataset detected by algorithms. Insights are made actionable by generating statistically accurate predictions about the relationship between the input and outcome variables, where the actions usually refer to simple decision rules (e.g., whether to classify an email as spam or not spam) (Dhar 2013; Clough et al. 2015; Bratteteig and Vern 2018). One consequence of this statistical approach to actionability in data is that the primary human learning comes from the data scientists working on given data on their computers. These data scientists operate at a distance from the settings in which phenomena occur or in which the data were collected or used. They process and explore the given dataset to develop models with desirable statistical properties. End-users (e.g., domain experts) who create and use data in their own settings, when they are made visible, only play a peripheral role of informing data scientists of domain-specific contexts as supplementary knowledge (Amershi et al. 2014).

A consequence of this focus on statistical properties has been that most research on data science tends to focus on models or algorithms rather than the data themselves (Hohman et al. 2020; Sambasivan et al. 2021). Data are already given. The key consideration becomes how suitable the data are to serve models and algorithms (e.g., how reliable or complete the dataset is) (e.g., Sun et al. 2017). Seeing data as being in service of algorithms (i.e., improving their performance) marginalizes the role of end-users who often initially collect and ultimately use the data. End-users are relegated to informers or to those who do the work surrounding the dataset (e.g., creating, collecting, managing data) that consists of tedious, uninsightful tasks, or the 'grunt work' or 'janitor work' of data science (Lohr 2014; Fletcher et al. 2020).

Finally, a consequence has been that the insights that could come from data become largely limited to statistical patterns in data and algorithmic predictions (Shmueli 2010). End-users' expectations of data science tools, however, go beyond making predictions. They include, for example, the ability to formulate a range of problems that they face in their work practices and settings in terms of relevant variables of data (Passi and Barocas 2019) and understanding and leveraging causal relationships between variables of data (Bhatt et al. 2020). The assumption that the primary insights from data are statistical predictions limits the relevance and value of data science tools to end-users' real-world practices and settings.

2.2.1.1 Mapping between inputs and outputs

By assuming that data are decontextualized and that insights from data can be reduced to statistical properties and predictions of algorithms, technical data science research tends to render largely invisible where data come from and the human work of making data actionable (e.g., Muller et al. 2019a; Aragon et al. 2022). The question of how data are selected and put together in this view is limited to a concern with stages of pre- and post-analysis such as how data are prepared and processed to serve statistical models and algorithms, and how the outcomes could be made to fit the phenomena that the data are assumed to represent (Offenhuber 2020). How data are selected and put together remains largely black-boxed. As Bowker (2013, p.170) puts it: "the interpretative work is done inside the computer and read out and acted on by humans…our interaction with the world and each other is being rendered epiphenomenal to these data-program-data cycles."

The invisibility of phenomena and humans during the process of analyzing data entails a view of data science largely in terms of mappings between input data and output data (Smith 2019; cf. Norman 1988). Mapping inputs to outputs need not involve human understanding, experience, or discretion about the actual phenomena (i.e., what it means to recognize a face). Smith (2019) uses the example of face recognition to illustrate this point: "the term "recognition" rather oversells what is going on. A better characterization is to say that [machine learning] systems learn mappings between (i) images of faces and (ii) names or other information associated with the people that the faces are faces of" (p.50).

Technical data science approaches often have little to say about the actionability of data as they relate to real-world phenomena. The relevance and significance of data is determined in terms of statistical patterns detected by models and algorithms (Khalid et al. 2014; Cai et al. 2018; Guyon and Elisseeff 2003). Statistical patterns in data may lead to novel patterns and insights for technical data scientists seeking to optimize the predictive accuracy of a model for a given dataset (Clough et al. 2015). It is less straightforward, however, how novel patterns in data, even if statistically valid, can be converted into practical action by end-users.

2.2.2 Making data actionable through human work surrounding models and algorithms

Several scholars argue that, to make data science models and algorithms understandable, data analysis techniques should reflect how humans think with data and not just reflect statistical techniques. Nadeem (2022), for example, stresses that, in addition to providing more information on models and algorithms, data science tools should focus on mirroring "human intuition" by providing "representations of statistics that best map to how humans intuitively think" (p.44). This human-centered view of data analysis in data science aligns with a core goal of research on end-user development (Paternò and Wulf 2017) of making computing processes more "natural" for developers, or "closer to the way that developers think" (Myers et al. 2017, p.2). Beyond just mapping statistical descriptive properties or correlations between input and output data, human-centered data science research focuses on how humans put together data cognitively. Tools for analyzing data should map not just input to output data, but map how humans think with data in their heads onto how representations are processed on computers (e.g., Green and Petre 1996).

Scholars of human-centered data science research and critical data studies emphasize that the actionability of the models and algorithms depends on making predictions not just accurate but understandable to humans (e.g., Aragon et al. 2022; Muller et al. 2019a). Issues with transparency, interpretability, and explainability reduce trust in models or algorithms by obscuring the sources of bias and arbitrariness (e.g., Rudin 2019; Lepri et al. 2018; Passi and Jackson 2018). Many data science techniques, especially the "blackboxed" data analysis processes of ML models, have been problematized for their lack of transparency regarding how a model and dataset combine to generate a prediction (Weller 2019).

Scholars in human-centered data science also emphasize that selecting and putting together data involves negotiations among stakeholders. Technical stakeholders (e.g., data scientists trained in statistical analytics) and non-technical stakeholders (e.g., domain experts or managers) are assumed to differ in their views and understandings of data. In their study of biomedical scientists collaborating with technical data scientists, for example, Mao et al. (2019) observe that building “common grounds,” or shared understandings and rules regarding what data to share and how, is critical to the success of data science projects. Based on ethnographic fieldwork of a data science team within a US corporation, Passi and Jackson (2018) show how data science projects involved processes for dealing with different goals, expectations, and skills between managers, business analysts, and data scientists to establish shared confidence and trust in the workings of data science models. Hou and Wang (2017) illustrate the importance of “brokers” who help facilitate communication, coordination, and knowledge sharing about data between domain experts and technical data scientists. These studies emphasize that designing and deploying data science systems is a multi-stakeholder endeavor that requires ongoing, situated work to negotiate what data mean in a given context (Passi and Sengers 2020).

2.2.2.1 Data scientists over end-users; models over data

Human-centered data science research still tends to focus on engaging with *given* datasets in relation to phenomena, and in ways still mostly separate from the real-world settings in which phenomena play out. Views of data and data science in this literature thus remain limited in accounting for the actionability of data in end-users’ practices and settings. Proposed theories and solutions still prioritize human intuition and cognition over the actual phenomena and environments in which humans are situated.

Studies in HCI and CSCW have only recently begun to explore the material situations in which data are collected and used (e.g., Stadelmann et al. 2018; Sambasivan et al. 2021). Beede et al. (2020), for example, provide an account of an AI-assisted eye-screen system deployed in clinics in Thailand. They show how the performance and effectiveness of the system was not determined by its predictive accuracy alone but depended on socio-environmental factors at the physical sites (e.g., poor lighting led to ungradable images; internet connectivity issues caused delays for patients). What remains little understood in extant literature is how end-users engage with and make decisions about their data in their own practices and settings.

While critiquing assumptions of data in technical approaches, human-centered data science research tends to focus on what can be done with models and algorithms, rather than what can be done with data to make data science more useful and actionable to end-users. The focus is on supporting data scientists or data science teams to more effectively deal with datasets for developing models and algorithms. Less has been said about supporting *end-users* to use and make decisions with data themselves in their practices and settings (Bhatt et al. 2020). Discussions of data have been subsumed under the extant focus on models and algorithms. For example, recent work on bias has focused on how biases in data lead to statistical biases in the models and algorithms, rather than a different framing of the domain itself. Smith (2019) observes, however, that the term "algorithmic bias" is potentially misleading: "It is the data…that is the primary locus of bias in machine learning results. The algorithms that run over the data are undoubtedly not innocent…[b]ut most examples of bias cited in the press and literature are due more to skewed data than to culpable algorithms" (p.67)

2.3 Proposals to make data and data science systems actionable

The different views of the data in data science systems described in 2.1 and 2.2 imply different understandings of what makes data and data science systems actionable. If data are viewed as decontextualized representations amenable to statistical analysis, the actionability of data and data science systems depends on the quality of the dataset used for developing models and algorithms (Polyzotis et al. 2017; Amershi et al. 2019b; Cabitza et al. 2019). If data are viewed as the result of human work, the actionability of data and data science systems depends on articulating, negotiating, and coordinating human values, intentions, and practices surrounding data collection and use; the goals of research and design are to offer "a practical, real-world guide for doing data science about people, with people, and for people in a way that is both ethical and responsible" (Aragon et al. 2022, p.147; also see Auernhammer 2020; Shneiderman 2020). I next provide an overview of existing proposals that reflect assumptions about data.

*2.3.1 Improving the quality of datasets (***post hoc***) for model work*

There is broad agreement that data in real-world data science activities is often unreliable (not accurately or consistently recorded (e.g., 'messy') and incomplete (e.g., missing variables or values such that there are not enough columns or rows of data for the model) (Muller et al. 2019b; Zhang et al. 2020; Sarker 2021). Data typically fail to meet basic standards of reliability and completeness for several reasons including the lack of documentation and guidelines for managing data (Redman 2018; Gebru et al. 2021; Mitchell et al. 2019; Sambasivan et al. 2021; Boyd 2021). Unreliable, incomplete data lead to time-consuming "data wrangling" tasks (e.g., curating, cleaning, formatting) and, more seriously, to "data cascades," or "compounding events causing negative, downstream effects" on model implementation (Sambasivan et al. 2021, p.5).

A variety of technical proposals have been advanced to support tasks of wrangling missing or messy data, and to ensure (in a statistical sense) the quality of the wrangled data. Proposals include developing interactive visualization methods to aid data transformation (Kandel et al. 2011), automating data inefficiency detection and repair (Hynes et al. 2017; Krishnan et al. 2016), incorporating the context of data's domain (e.g., master data, reference data) into wrangling tools (Koehler et al. 2017), and better aligning data wrangling tools and data scientists' data exploration workflows (Drosos et al. 2020).

These proposals for making the work of wrangling data more efficient, however, are mostly *post hoc*. The proposals concern data work that is supposed to be done mostly *after* the data have been collected (e.g., Pine and Liboiron 2015). “Data work” may be done by humans, but it is still separate from the settings in which phenomena happen and in which humans (i.e., end-users) in these settings seek to engage with phenomena through data. The relationship between data and phenomena is still assumed to be stable, and the technical work of data science projects is typically assumed to be the ‘one-off application’ of a statistical model to a given static dataset (Polyzotis et al. 2017). Changes to data (or its underlying distribution), called ‘data drift,’ are detrimental to model performance and to be controlled (Hohman et al. 2020; Hoens et al. 2012; Williamson and Henderson 2021). While several studies explore ways to set up a ‘pipeline’ for data science activities with continuously incoming, and possibly changing, datasets (e.g., Roh et al. 2019; Breck et al. 2019; Lourenço et al. 2019), these activities still mostly center on data scientists working on their computers.

These proposed design solutions pay less attention to the relationships between phenomena and humans through data, and have limited relevance to the goal of this dissertation to explore how data science systems can be more actionable for their end-users in their practices and settings.

2.3.2 Understandable models and algorithms

In human-centered data science research, data science systems need to be understandable to humans to be actionable. Scholars have sought, for example, to increase the interpretability and explainability of models and algorithms in data science systems by making data visualization and analytics tools more user-friendly and interactive, such as by designing tools to enable data science systems users to experiment with statistical relationships between input variables, models, and their outcomes (e.g., Mitchell et al. 2019; Krause et al. 2016; Amershi et al. 2014). Several scholars have been interested in providing support for non-technical end-users (e.g., domain experts). Amershi et al.’s (2014) “interactive machine learning” framework seeks to enable end-users to iteratively experiment with machine learning tools with minimal technical expertise. Gil et al. (2019) propose “human-guided machine learning” systems to enable end-users to prepare data and learn with different data models by themselves while leveraging their domain knowledge.

On the interpretability of data science models, Rudin and Radin (2019) argue that we should try to do away with the assumption that ML models are inevitably opaque. They present examples in which there was no meaningful difference in predictive accuracy between ML models with little interpretability using a large number of variables and much simpler, inherently interpretable models using a small number of variables that are relevant to a phenomenon of interest (e.g., a recidivism prediction model based on only age and criminal history) (Angelino et al. 2018; Zeng et al. 2016).

Recently, documentation about datasets and algorithms has also emerged as a potential solution to make data and data science systems more actionable. Proposed solutions include “Model Cards” that document general characteristics of ML models (e.g., how ML models were trained, how to measure their performance) (Mitchell et al. 2019), “FactSheets” that document similar characteristics about AI services (Arnold et al. 2019), and “Data Readiness Reports” that include automatically generated information on the characteristics of datasets, the quality of datasets used, recommendations for addressing data quality issues as well as restrictions and policies on data usage (Afzal et al. 2021).

2.3.3 Making human work visible

Another strand of human-centered data science research explores the work processes of “data science workers,” a term meant to encompass data scientists and all other people for whom data-related activities are central to their work. Muller et al. (2019) conducted an interview study to analyze the workflows of data science workers, identifying five dimensions of human interventions: data as given, as captured, as curated, as designed, and as created. Building on Muller et al.'s (2019) work, Zhang et al. (2020) identified how data science workers collaborate across different stages of a data science workflow.

Other scholars have focused on extending quantitative methodologies to account for human knowledge and reflexive activities in engagement with data. Tanweer et al. (2021) propose a systematic application of qualitative critical methods, specifically interpretivism, abductive reasoning, and reflexivity to make data science research "more reliable, more thorough, and more ethical." Marres (2020) seeks to introduce into data science the interpretive method of situational analysis (Clarke et al. 2015) that takes a situation involving various human actors, non-human actors, and other organizational and technological elements as a unit of analysis.

Several studies have also begun to explore methodologies to increase human participation in the design of data science projects (e.g., Lee et al. 2019; Friedman and Hendry 2019). Bratteteig and Verne (2018) examined challenges of using participatory design methods in AI projects. They suggested a need to support participants to understand what data science techniques can and cannot do and imagine and evaluate possible activities. Liao and Muller (2019) proposed a method for data science and AI systems that combines participatory design and design fictions to explore stakeholders' values and speculate about future scenarios. Sloane et al. (2020), however, identified different types of participation and argued that not all types would empower humans. Participation through data work (e.g., content moderation, mTurk workers, reCAPTCHA), for example, should require participants' consent and they should be offered options to opt out and proper compensation.

Human-centered data science research implies that the actionability of data depends on how humans work with data, but so far mostly with respect to how data affects the performance of models and algorithms. This research still puts technical data science work (e.g., of data scientists and data science workers) at the center of data science systems. There has been less attention to how end-users of data science systems use their data in relation to their phenomena and artifacts of interest in their material practices and settings. In my dissertation, I consider how end-users' data practices can be central to making data science systems more actionable. To set up this point of view, I next sketch out what I refer to as a phenomenological view of data.

2.4 A phenomenological view of data: making central end-users' data practices

The views of data in recent data science research (2.1-2.3) are in some sense diverse. Technical data science research tends to view data as objective, abstract representations of phenomena transformed by models and algorithms into statistical predictions through largely blackboxed processes of analysis. Human-centered data science research views data as representations contingently associated with phenomena, and where the collection, use, and analysis of data are reflective of human practices, values, intentions, and systems. Despite such differences, existing approaches to data in data science research commonly assume that the data in datasets are already collected and that the central actors who engage with data are technical data scientists. As Gil (2017) argues, the use of so-called "intelligent" data science and AI systems has been reduced to data scientists' application of a mostly given set of techniques to well-established tasks and phenomena. These existing approaches make it challenging to imagine how data science systems could serve end-users across different real-world practices and settings.

To explore how data science systems might become more actionable for data users, I propose turning attention to how end-users (e.g., domain experts) might engage with data science systems in their material practices and settings. I seek to go beyond how domain experts may inform or collaborate with technical data scientists. My proposal is to frame end-users of data science systems as types of *data users* - those who collect and use their data as part of their daily practices and settings - and to imagine their data practices as a foundational topic in data science research.

As a basis for this argument for actionable data science systems driven by end-users' data practices, I develop an approach to data that is informed by phenomenological views of humans' interaction with technology (e.g., Agre 1997; Smith 1998; Bowker 2013). Set against dualist or purely mental constructions of the world, phenomenology focuses on how humans' relationship and engagement with material things and environments shape their being in the world (Olsen 2010). By examining processes through which humans experience the world with things in a situated manner, phenomenology concerns "questions of how to act in everyday situations and relations" or "possibilities of creating formative relations between being and acting, between who we are and how we act" (Van Manen 2007, p.13). Drawing on this focus on the situated, material experience of humans with pragmatic concerns of what can be done and how, I next unpack how a phenomenological view opens up a concept of data that helps account for the data practices of end-users situated in their own settings and practices.

2.4.1 Relating data to phenomena: data as presented to data users in their material settings

In this section, I seek to frame a phenomenological view of data to inform the approach that I will propose for making data science systems more actionable for end-users. To set up my framing, I first contrast implicit understandings of data as *representations* versus a more phenomenological view of data as *presentations*. In the data science research described in 2.1-2.3, data tend to be implicitly viewed as *representations* of phenomena. By representations, I mean that the symbols are assumed to be detachable from material phenomena, and manipulable by humans or computers in the absence of the actual phenomena (Smith 1998, p.287; Clough et al. 2015;

Thrift 2008). This assumption of separation and absence makes it possible to study the relationship between data and phenomena in the abstract and to scale and generalize data analysis across many contexts. Yet this assumption of data as detachable from phenomena makes it difficult to examine and deal with the world the way it really is in any particular context (Smith 2019).

In a more phenomenological view of data, the focus shifts from "re"-presenting phenomena as data after the fact to experiencing the *present* phenomena through and with data (Offenhuber 2020, p.99). In this view, the standard dictum in information visualization of "[a]bove all else show the data" (Tufte 2001, p.105) becomes "above all else move data closer to the phenomenon" (Offenhuber 2020, p.104). When we move data closer to the phenomenon, the relationship between data and phenomena is no longer just a correspondence or mapping (e.g., how what is captured by data corresponds or maps to phenomena out there). In a phenomenological approach, data as well as data users are always situated and cannot be separated or detached from phenomena that they are experiencing. Data, phenomena, and data users are entangled and oriented to each other through rich, complex, and messy interactions in the real world (Olsen 2010; Matthews 2002; Dreyfus 1992; Auernhammer 2020).

In a phenomenological approach to studying the actionability of data, data are not given as an abstract dataset. Data *give themselves* to situated data users as something that already relates to, or is part of, a phenomenon that is present (Merleau-Ponty 1962; Smith 1998). Data are presented to, and experienced differently by, data users depending on their domain of practice. To foreshadow the findings of chapters 3 and 4, for a brewer, the temperatures of the brew in a fermentation tank are not abstract numbers but data that tell the brewer something about the present brew's state and character given the particular settings of the brewery. The givenness of data here refers to how data *present*, rather than re-present, themselves to situated humans to be experienced by them (Marion 2002; Dreyfus 2007). In their presentation to data users engaged in practices in material settings, data about phenomena appear to seek a response (e.g., the brewer adding a chemical to the brew based on the temperature) (Dreyfus 2002), rather than showing themselves as neutral, objective resources amenable to abstract, statistical manipulations.

2.4.2 How data become actionable: registering and familiarizing with "things"

In a phenomenological view of data and data science systems, neither things in the world (e.g., a brew being fermented) nor data about them (e.g., recorded variables such as temperature and pH as well as unrecorded variables such as smell and appearance) could be simply assumed as given. Data are presented and experienced in ways that depend on the material situation and relate to their material properties (e.g., of the particular batch, having a particular color or flavor). How do we experience things with or through data?

Smith (2019; 1998, p.282) stresses that things in the world need to be "registered," or made "ontologically intelligible" or accessible to support humans' intended activities (e.g., a brewer makes sense of a brew being fermented through temperature and pH data). Registration involves delineating boundaries, assigning identities, and focusing on particular properties of things while ignoring others to make sense of incredibly rich and complex things in the world; it is to "carve the world in such a way that violet is closer to blue than to yellow, that rutabagas are closer to parsnips than to blueberries, that melancholy is closer to ennui than to spunk" (Smith 1998, p.280). The process of registering things through data, however, is never fixed, and is subject to destabilization and restabilization. As the situatedness of humans and things changes, so does the appropriateness of what has been registered and how. The purpose of registration is to "make the world present and to be present in the world" according to a given situation (Smith 1998, p.306; Smith 2019).

Processes of registration need to be accountable not just in the sense that the particular choices that humans make need to be made visible (e.g., documenting decisions or properties of datasets or algorithms as in human-centered data science research). In a phenomenological view, processes of registration need to "support, rather than [undermine], the world's status as world" (Smith 2019, p.102) and are only valid "*if they make sense of the world as world*" (p.103, emphasis original). To continue with the brewing example, a phenomenological view of data implies that the brewers' engagement with data that are presented should enable them to make sense of and work with the brew as it really is in their situation. Statistical analysis is potentially in support of, but not fundamental to this engagement. How the brewer selects properties of data out of all possible or available data (e.g., focusing on smell and temperature rather than pH) depends on their material settings and practices (e.g., cold, dry weather and tight production schedules), and these selections of data drive the effects of data on practice more than statistical analysis.

In a phenomenological view of data, data are made actionable as they help data users become familiarized with the situations that they are thrown into and the things that they deal with in those situations so that they could *go on* (Heidegger 1962; Dreyfus 2002; Dreyfus 2007). Artifacts emerge as a critical element in data users' engagement with the world through data; data presented to their users are always about or of artifacts (e.g., the temperature or smell is of a brew being fermented, telling something about the brew). As Borgmann (2015, p.249) illustrates: "[w]ithout some stable and identifiable thing at the center, variants would be different independent entities, and the multistability of interpretations would turn into a multiplicity of objects."

2.4.3 Registering properties of things through data

How do data users register things and artifacts in the world through data? A phenomenological approach to data allows us to explore this question in relation to the role of things or artifacts in situated data use. Take Heidegger's (1962) famous distinction between artifacts' ready-to-handness and present-at-handness. When we are in ready-to-hand mode, a thing is transparent and unobstructive, and we unselfconsciously deal with it (e.g., reading a document on the computer screen). When the thing becomes problematic (e.g., black horizontal lines appear on the screen), however, we become more aware of the now-less-transparent thing and deliberately try to see what is going on and deal with it; we then move to a present-at-hand mode. Rather than viewing the two modes as different kinds, in most purposeful activities we do, our engagement with things could be understood in terms of the "gradation of "the coming to mind"" (Olsen 2010, p.164) involving how we flexibly maintain or modify things as we continue to engage with them to improve the given situation (Dreyfus 2007; Agre 1997).

In a phenomenological view of data, data users' engagement with things can therefore be understood as mediated by *data about the properties of these things* of different degrees of present-at-handness (e.g., how black horizontal lines or other properties of the screen appear to an end-user and what they can do next to make reading on the screen less problematic). Data users engage with data about the properties of things that "stan[d] forth" (Heidegger 1971, p.166) to them. The idea of "standing forth" suggests that things have properties. When we make sense of or engage with things, only a particular set of properties of things present themselves, while other properties remain concealed. We always "partially, perspectivally" engage with things (Van Manen 2014, p.62) through data about a particular set of properties.

The selection of data depends on the relevance and significance of properties of things in a given situation (Dreyfus 1992). Properties of a present-at-hand thing relevant to an intended activity "light up" and "come to mind" to a data user. To understand these properties, the user would need to step back from the situation and deal with data about these properties in a more abstract way, but doing so presupposes that properties of data were already *presented* to the user (e.g., how the properties connect to each other and to our intended activities) (Heidegger 1962; Olsen 2010). These properties of data require responses by data users (i.e., taking actions) specific to their material situations. Since there could be multiple possible responses or a response tried out may fail to improve the situation, the users of data need to be able to modify or refine their responses in an iterative and dialectic manner (Dreyfus 2002; Brown 2011).

When data about the relevant properties of artifacts and things are put together, properties of data are seen as neither completely random nor pre-defined patterns amenable to statistical analysis. Properties are also not simply a product of social, discursive, or symbolic relationships. Properties *materially* mediate relations within and surrounding the thing itself. Properties and their relations are held together as a unified thing. This unity is not "atomic" but has some coherent "internal structure" (Smith 1998, p.230, pp.278-282). For the actionability of data and data science systems for end-users, the relevance and significance of properties of things depends on the coherence of the conceptual and material constraints among these properties. For example, a brew or a photo holds together material properties and relations, be they microbiological or visual, in a coherent manner that makes it understandable and engageable as *this* particular *brew* or *that* particular *photo*. Olsen (2010) argues: "rather than thinking of these different properties as produced in relations, we may think of them *as what makes relations possible*" (p.158, emphasis original).

2.4.4 A phenomenological view of data and data science systems

A phenomenological view of data implies that data are presented as relevant to properties of artifacts and things that stand forth in data users' practices and settings. As users of data, end-users are the ones who select and put together data about the relevant properties. They draw on their understanding of conceptually and materially coherent structures of properties to flexibly and skillfully respond to artifacts and things in given situations. A phenomenological view of data and data science can therefore be summarized as: (i) data are *presented* to end-users of data science systems, (ii) these users use data to *register* properties of artifacts and make them intelligible, and (iii) the processes of registration involve putting data about properties of artifacts together in ways that are coherent with their conceptual and material constraints.

In a phenomenological view, data are not simply objects to analyze to uncover some patterns. Data are coherent sets of properties that allows data users to understand and participate in the things around them (e.g., Olsen 2010). In Smith's (1998) terms, the objective of a phenomenological approach is to develop "an account of *participation* or *engagement*…not just of cognition, and not even just of experience" (p.287, emphases original). In this approach, actionability concerns how data users could efficiently and effectively select and put together data about coherent sets of properties of things in their settings to direct practical action. Attention shifts from how to conduct or support statistical analysis, where data are understood in terms of models and algorithms (e.g., their inputs and outputs), towards determining what to do with data and how to use data to participate in and engage with things in their material settings.

What are the implications of a phenomenological approach to data for making data science systems more actionable? Recent research on data science systems consistent with a phenomenological approach implies that a fundamental issue for actionability is how a data science system mediates the *phenomenological fit* between data and artifacts (Sambasivan et al. 2021). Freiesleben et al. (2022), in a similar vein, propose that we make data science phenomenon-centric rather than model-centric to direct research efforts from describing and explaining models to learning about real-world properties of phenomena. If data and models do not properly approximate the phenomena, they argue, any conclusions drawn from the data and models will have limited value and relevance to end-users. If end-users want to use data to properly participate in and engage with artifacts or phenomena in their settings, it becomes critical to consider how well the design of a data science system supports users' ability to coherently capture relevant properties of artifacts and phenomena.

Pietsch (2022, *passim)* argues for viewing data science as a "phenomenological science," meaning that data are used for participatory engagement with real-world artifacts and phenomena, not just to develop a general theory about them. The purpose of data science is "difference-making" (*ibid., passim*), that is, making a change in a phenomenon of interest in a given situation to manipulate it as intended. The knowledge produced in data science is always "local," "contextual," and "approximate" in that it is not possible to account for a situation "in the wild" fully in advance (p.40). He contrasts this with the aim of a theoretical science, which is to produce an abstract, precise, and complete explanation of phenomena in the world that could be universally applied across different contexts. That is, data science as a phenomenological science seeks an approximate understanding of an artifact or phenomenon to a level that is sufficient for end-users to make it effectively actionable in a given situation. The implication is that, for a data science system to be more actionable, it should aim not for optimizing the accuracy of predictions, but for enabling end-users to develop coherent sets of properties that approximate phenomena.

There are, of course, many contexts where finding statistical patterns in data is meaningful in itself, such as studying theoretical astrophysics or a huge population (e.g., Dreyfus 1992; Leonelli 2015). A phenomenological approach to data, however, draws attention to processes by which humans as data users engage with data to take actions on things and phenomena in their material settings.

This dissertation seeks to develop a phenomenological approach to data and data science systems empirically by describing processes by which data users, as potential end-users of data science systems, experience and register data as properties of things situated in their material practices and settings. It will draw on several key concepts that I define for the purpose of this dissertation as follows:

- Data: *properties of artifacts and things presented to a human (end-user) to help form the basis of their engagement with phenomena*
- Data users / end-users: *humans who collect data about artifacts and phenomena of interest in their practices and settings and who use their data to engage with and act on the artifacts and phenomena; data users are also understood as end-users of data science systems.*
- Artifacts / phenomenon: *(i) something that appears to be, or is presented to us as having multiple properties but is held together as an identifiable and coherent whole (Redström and Wiltse 2018; Harman 2011), (ii) something on which "one can have a perspective" (e.g., one can make sense of a brew as a thing in a particular way) (Smith 1998, p.117), and (iii) something that serves a purpose and has an effect on the humans and settings involved (Heersmink 2021, p.3).*

Based on these concepts, the following chapters will analyze four studies of data practices across three different settings: craft brewers trying to consistently produce high-quality and sellable beers at a brewery (chapter 3), a pilot ML project at the brewery (chapter 4), people with visual impairments trying to use their personal photos on their smartphones (chapter 5), and repair workers trying to fix broken machines (chapter 6). To develop an end-user approach to making data science systems more actionable, the dissertation will seek to surface and unpack commonalities across the settings through which these *data users* interacted with data about artifacts and things to engage with phenomena in their practices and settings.

3 How Domain Experts Work with Data

In this first study[1], I explore how craft brewers, as domain experts, use and work with their data in their practices and settings. Domain experts play an essential role in data science by helping data scientists situate their technical work beyond the statistical analysis of large data sets. How domain experts themselves may engage with data science tools as a type of end-user remains largely invisible. Drawing on an ethnographic study of a craft brewery in Korea, I show how craft brewers worked *with* data by situating otherwise abstract data within their brewing practices and settings. The study contributes theoretical insight into how domain experts use data distinctly from technical data science in terms of their view of data (situated vs. abstract), purposes for engaging with data (guiding processes over accurately predicting outcomes), and overall goals of using data (flexible control vs. precision). I then discuss four ways in which working *with* data can be supported through the design of data science tools, as well as how craftwork can be a useful lens for integrating domain expert-driven understandings of data science into CSCW and HCI research.

3.1 Introduction

A long-standing focus in CSCW has been studying how processes of designing and using technological artifacts are situated in domain experts' work practices and settings (e.g., Suchman 1983; Orr 1996; Dittrich et al. 2009; Ackerman 2000]. This focus has led to research on supporting domain experts to bring their experience and points of view to bear on the design and use of

[1] *This chapter has been published as Jung et al. (2022a). I proposed and conducted the ethnographic study. I analyzed the findings and wrote the manuscript with feedback from the co-authors (Tom Steinberger, John L. King, and Mark S. Ackerman).*

technological artifacts, such as through collaboration between software developers and domain experts as end-users, or among domain experts who develop software for themselves (e.g., Mørch and Mehandjiev 2000; Howison and Herbsleb 2011; Segal 2009; Jirotka et al. 2013).

In the emerging research on data science in CSCW and HCI, several scholars similarly aim to support the involvement of domain experts in the context of designing and using data science tools (Amershi et al. 2014, Muller et al. 2019b; Seidelin et al. 2020; Gil et al. 2019). This research observes that technical data scientists and nontechnical domain experts have different goals and points of view in data science projects (Mao et al. 2019; Passi and Jackson 2018). Collaboration depends on technological and organizational solutions to increase domain experts' involvement in data science activities by, for example, making it easier for them to manipulate data sets or share knowledge with technical data scientists (Andrzejewski et al. 2009; Amershi et al. 2019a; Park et al. 2021).

The role of domain experts in this research, however, tends to be confined to informing technical data scientists of their domain-specific knowledge so that the technical data scientists are able to make sense of data sets and build models that provide relevant and actionable insights (Ribes et al. 2019). What is still largely invisible in the emerging CSCW and HCI research on data science is how domain experts themselves may engage with data science tools as a type of end-user. Given that data relevant to experts' domains are situated in their work practices and settings (e.g., Taylor et al. 2015; Feinberg 2017; Passi and Jackson 2017; Koesten et al. 2021; also Suchman 1983; Lave 1984), the success of a data science project can be helped by understanding problems for which domain experts use data (Mao et al. 2019; Passi and Jackson 2018).

To understand how domain experts use data, this paper builds on an ethnographic study of a team of brewers at a craft brewery in Korea who produced and used data to improve the quality and consistency of their beers. Craftwork such as craft brewing is data-intensive and requires deep domain expertise (Sennett 2008); craftwork also has served as a sensitizing concept for understanding the human side of data science activities (Muller et al. 2019a; Muller et al. 2019b; Raman and Hellerstein 2001) and generally in CSCW for examining the materiality of work practices (Rosner 2012; Lingel and Regan 2014; Cheatle and Jackson 2015). A craft brewery thus offers rich possibilities for understanding how domain experts use data in their work practices and settings.

We present three vignettes that surface the processes by which craft brewers situated their data in their brewing practices and settings to deal with changes in the data that arose as materials and circumstances continually shifted. We characterize these processes as working *with* data, to refer to how the brewers used data by situating it *within* their craftwork, and not just by applying domain knowledge to technical data science work (e.g., wrangling data sets or interpreting the predictions of machine learning models).

A main implication of our findings is that the brewers' situating of data in their brewing practices and settings was often at odds with common understandings of data science as the statistical analysis of large data sets. Data science tools (including most machine learning models) typically require assuming that data come from the same statistical distribution (Bengio et al. 2021, p.63), whereas the shifting materials and circumstances surrounding the craft brewers' data made this assumption unreliable.

We discuss how this main implication of our findings leads to both theoretical and practical contributions. First, we show that a focus on how craftworkers used data as a type of domain expert leads to a distinct understanding of data science in terms of assumptions about the nature of the data, the purposes for engaging with data, and the overall goals for using data. Second, we discuss how a focus on working *with* data points to distinct challenges for the design of data science tools. By highlighting how and why data science for craftwork can be theorized and designed for, our study contributes towards integrating domain expert-driven understandings of data science in CSCW and HCI (e.g., Banovic et al. 2019; Gil et al. 2019). Our paper also contributes by extending the CSCW literature on craftwork to how craft workers work with data.

3.2 Background

In this section, we survey prior research in CSCW, HCI, and the social sciences on (i) the role of domain experts in data science activities and (ii) craftwork as a lens for studying domain experts' data science activities. We then outline how craft brewing can be a useful site for exploring the connection between data science and craftwork.

3.2.1 Domain experts in data science

Data science is commonly understood as applying statistical analytics techniques such as machine learning models to big data. Much research has examined how technical data scientists — i.e., data scientists professionally trained in statistical analytics techniques — make decisions (Kery et al. 2019), perceive and understand automated tools and their results (Wang et al. 2019b; Hohman et al. 2020; Kaur et al. 2020; Drozdal et al. 2020; Xin et al. 2021), and collaborate with other technical data scientists (Wang 2019a; Piorkowski et al. 2021; Zhang et al. 2020). Yet there is no agreed-upon definition of a data scientist (Miller et al. 2019a; Carter and Sholler 2016). CSCW research characterizes a data scientist as anyone for whom data-related activities are central to their work, meaning that the work depends on the ability to build and operate data management and analytics tools (Boukhelifa et al. 2017).

Several scholars in CSCW and HCI have begun to study how nontechnical stakeholders engage with data in data science projects. Muller et al. (2019b) propose expanding the term "data scientist" to "data science worker" to refer to a wider range of data practices than statistical analytics techniques. Boukhelifa et al. (2017) and Liu et al. (2020, p.66) examine "data workers" as domain experts who "perform data analysis as part of their work but would not call themselves technical data scientists." In their study of biomedical scientists collaborating with technical data scientists, Mao et al. (2019) observe that building "common grounds," or shared understandings and rules regarding what to share and how, is critical to the success of data science projects. Hou and Wang (2017) and Choi and Tausczik (2017) illustrate the importance of "brokers" who help facilitate communication, coordination, and knowledge sharing between domain experts and technical data scientists. Taken together, extant research in CSCW and HCI has studied domain experts' involvement in data science activities primarily in terms of applying their domain knowledge to inform the work of technical data scientists (Carter and Sholler 2016; Zhang et al. 2020; Frey et al. 2020; Park et al. 2021; Viaene 2013).

As data management and analytics tools have become more user-friendly and accessible, a number of studies are emerging that consider domain experts more as end-users who could directly engage in technical data science activities (Seymoens et al. 2018; Gil et al. 2019; Barczewski et al. 2020). Amershi et al.'s (2014) "interactive machine learning" framework seeks to enable domain experts to iteratively experiment with machine learning tools with minimal technical expertise. Schobel et al. (2017) offer a framework that gives domain experts the capability to transform and design data collection tools for data analysis. Gil et al. (2019) propose "human-guided machine learning" systems to enable domain experts to prepare data and learn with different data models by themselves while leveraging their domain knowledge. These studies are consistent with a view of domain experts as “owners of problems” who could “act as designers” of technical data science activities themselves (Fischer et al. 2009, p.37).

Prior CSCW research suggests that such a view of domain experts’ involvement in data science activities depends on viewing data as situated in their work practices and settings (e.g., Karasti et al. 2006; Kitchin 2014a; Passi and Jackson 2017; Passi and Jackson 2018; Feinberg 2017; also Orr 1996; Suchman 1983). Borgman et al. (2012) examine a scientific infrastructure project involving environmental scientists, computer scientists, and engineers, and find that their collaboration depended on negotiating different perspectives on what constitutes data. Based on their ethnographic study of the data-related activities of an urban street community, Taylor et al. (2015, p.2864, emphasis original) propose the term “data-in-place” to “think of [data] in terms of a *social geography* in which data, people, and things intermingle to continuously enact place.” Supporting domain experts’ involvement in data science activities could be informed by further insight into their situated use of data.

3.2.2 Craftwork as a lens for studying domain experts’ use of data

Craftwork concerns the process of producing an artifact that is consistent in quality, has some utility for someone else (i.e., a customer), and which is based on skillful engagement with unpredictable materials and settings (Kritzer 2007; Becker 1978). A craft worker is distinguished by “an ability to handle a wide variety of techniques and materials with ease and efficiency” (Becker 1978, p.888) to creatively respond to contingencies and achieve consistent quality (Bardzell et al. 2012). In CSCW and HCI research, craftwork has served as both a metaphor and empirical site to examine situatedness in the work practices of software development (Lingel and Regan 2014), engagement with the physical and digital materialities of artifacts (Golsteijn et al. 2014; Cheatle and Jackson 2015), critical thinking and making (Ratto 2011; Lindtner 2014), the notion of quality in design (Bardzell et al. 2012), and the data practices of data science workers (Muller et al. 2019b).

While not all domain experts are craftworkers in the sense of the paragraph above, craftwork offers a valuable lens for linking domain experts' work practices and settings to creative technical work, including data science activities. Sennett (2008) finds a close affinity between craftwork and scientific or technical computing work in their basis on an oscillation between "problem finding" and "problem solving," such that "the ancient potter and the modern programmer [are] members of the same tribe" (p.26) (also see (Raman and Hellerstein 2001) for use of this metaphor in regard to data science activities).

Consistent with the empirical orientation of domain experts in data science, craftwork involves not just "thinking *about*" technical aspects of a static problem, but iteratively "thinking *with*" changing tools, materials, actors, and environments to solve a problem as it evolves (Marchand 2016, p.12, emphases original; Sennett 2008). "Thinking with" data in craftwork points to the inherent tensions over formulating and translating problems in data science projects (Passi and Barocas 2019). In their study of collaboration between biomedical scientists and data scientists, Mao et al. (2019) illustrate that the ability to understand and use medical data depended on knowledge of "medical practices, clinical trial routines, regulations and direct impacts on patients [that do] not come with the meta-data or protocols" (p.14) and the lack of the ability to "think with" data often resulted in asking abstract questions whose relevance to a given domain was uncertain. Marchand (2016) finds that iteratively "thinking with" various types of data is integral to how craftspeople engage in "active and regular seeking of irregularities, mistakes, deficiencies, and inefficiencies" (p.16) to gain a sense of mastery over their unpredictable materials and contingent work practices.

3.2.3 *Craft brewing and data science*

Craft brewing is distinguished from industrial brewing in the smaller size of the breweries, the prioritization of human skills over automation, and an ability to develop various beer styles that leverage unique raw materials in addition to more standardized core offerings (Morgan et al. 2020). The few studies on craft brewing in CSCW and HCI focus on support for communities of amateur brewers or producer-consumer engagement and concern research on social computing (e.g., Kuznetsov et al. 2016; Knearem et al. 2019a; Knearem et al. 2019b). Our study instead focuses on craft breweries' work practices for producing small batches of beer based on leveraging unique raw materials controlled by a small number of brewers (Morgan et al. 2020; Lewis and Young 2001).

Craft brewing is an example of craftwork salient to studying domain experts' use of data. First, most craft breweries still rely on domain experts (the brewers) engaging with conventional equipment and techniques, and are just now beginning to explore advanced data science tools to engage with brewing data (Fisher et al. 2020). For example, machine learning algorithms for analyzing various data on the brewing process have emerged for predicting alcohol concentration during fermentation (Fisher et al. 2020), automating the evaluation of beer quality parameters (Lukinac et al. 2019), and evaluating beer acceptability (Viejo et al. 2019).

Second, as an empirical site, craft brewers' production and use of data is critical to quality control and consistency in their brewing practices (Fisher et al. 2020; Lima et al. 2011). The grain-to-glass brewing process makes use of various types of data such as grain inventories, recipes, and temperature, pH, alcohol by volume (ABV — the percent of a volume of liquid that is alcohol), and carbonation readings that continually need to be measured and interpreted (Morgan et al. 2020; Bamforth 2003). To achieve consistent quality, craft brewers learn to engage with the particular specialty brewing ingredients (e.g., malts, hops, yeasts) and equipment (e.g., grain mills, lauter tuns, fermenters, hydrometers) used in their brewhouse (Kritzer 2007). Evaluating the quality of beer is not straightforward. Even when parameters and specifications are precisely defined, accurately and efficiently measuring and monitoring brews poses an organizational challenge, especially for small craft breweries with limited resources (Lewis and Young 2001).

Brewing is also a delicate microbiological and biochemical process (Hough 1991; Lima et al. 2011). As Bamforth (2003, p.86) explains: "a vast myriad of compounds and physical interactions influence the quality of beer. Let just one of them be out of balance, and the whole product will be ruined." The relationships among this vast number of variables are difficult to identify or isolate, posing challenges for the brewers to take corrective actions by removing or adjusting individual variables once the brewing process has begun. Hughes (2009, p.61) stresses that ensuring consistency in producing a beer recipe calls for "a fine degree of control" by closely attending to the patterns and changes in data throughout the whole brewing process.

Studying how craft brewers use data could thus point to distinct domain expert-driven understandings of data science tools and their design. How do craft brewers, as domain experts, work with their data to find and solve problems as their materials and circumstances continually shift? What are the challenges that arise in situating their data in their work practices and settings? And how could craft brewers' use of data inform the design of data science tools for domain experts more generally? This paper seeks to explore these questions in our study of a craft brewery in Korea.

3.3 Data and methods

3.3.1 TheBrewery

Our study builds on ongoing fieldwork at a four-year-old craft brewery, called TheBrewery, located in Korea. The brewery was run by four male brewers, who all started as homebrewers, and one male intern (see **Table 1**). It produced some of the most popular craft beers in the region, securing nationwide distribution in convenience stores and supermarkets despite its small scale. Korea's craft beer scene took off only in the early 2010s, led by the expat community who had earlier exposure to and knowledge of craft brewing. The original head brewer at TheBrewery was also an expat and had several years of brewing experience both in his home country and Korea. Using information available online and gathered from his prior work, he created the basic brewing processes, beer recipes, and brew sheets. At the time of the study, the head brewer had recently left the brewery and these materials were still in use with minor modifications. While the facility's longest-serving brewer (B1) acted as an informal leader, all brewers had similar levels of brewing experience and there was no official head brewer. The brewing team also collaborated with the staff at the brew pub in the same region as well as the marketing team, the distribution team, and the management team (who were stationed in Seoul, several hours' drive from the brewery, and who regularly visited the facility), communicating with them via email, phone, and messaging apps (e.g., Slack) as well as in person.

The brewing team was responsible for everything that happened at the facility, including inventory management, production, and facility maintenance. The brewers determined the production schedule themselves based on their communication with other teams. The schedule was subject to frequent changes due to reasons including supply issues, fluctuations in demand, external inspections or events, or just balancing workloads. The schedule was manually recorded on a whiteboard calendar in their office/meeting room at the facility.

The brewers used a three-batches-in-one-brew system, meaning that they would make three batches and then combine these batches in one fermenter (the reason being that the fermenters were much larger in volume than the brewing equipment, as is typical in a brewery). When a brew would be scheduled, the three responsible brewers would get together the previous day to examine the existing recipe, the previous brew sheet data, and current inventory levels to decide if they wanted to make any adjustments to the recipe or the brewing process. Brewing each batch usually took around half a day and three batches were brewed consecutively over two to three days. Each brewer recorded measurements and comments on the brew sheets throughout their shift, and at each handoff point, the two brewers inspected the brew sheet data to communicate any information that the next brewer needed to take note of (e.g., the status of the grain mill on that day) to remedy problems detected in the previous batch (e.g., if pH is too high, the next brewer should add more water). The three batches would each be transferred to a single fermentation tank. The fermentation process then was monitored and recorded on a separate fermentation log. Once fermentation was complete, the brew was transferred to and stored in a bright tank, or a tank for holding fermented beer prior to packaging. Based on their distribution schedule, the brewers would package the brew either in bottles or kegs, recording data on a packaging log.

In addition to this production-to-packaging process, the brewers purchased ingredients, reorganized the grain storage room, cleaned up the facility, repaired broken equipment, held informal discussions based primarily on their brew sheets, performed brewing-related research online individually and as a team, took snack breaks, or just had fun being together in the idyllic location of the brewery on the east coast of Korea.

Table 1. Members of the brewing team at TheBrewery

Brewer (B) Head brewer (HB) Intern (I)	Age	Nationality	Experience at the brewery	Notes
B1	30s	Korean	3.5 years	Takes an informal leadership role
B2	30s	Korean	3 years	
B3	30s	Korean	3 years	
B4	40s	American	2 years	
HB1	30s	Australian	3 years	Former head brewer; left the facility the previous year
I1	20s	Korean	4 months	Monitors and logs the fermentation process; cleans the facility

3.3.2 *The data problem*

While still accounting for only three per cent of the country's entire beer market, the craft beer sector has been showing strong growth in Korea (Seo 2018; Kim 2020). Well-established in the region, TheBrewery's management team at the time of our study was seeking to grow and scale up operations. They were interested in ways of enhancing production efficiency and improving the brewers' skills, particularly through better utilization of their brewing data that were being recorded but not systematically analyzed.

The data central to the brewery's brewing process came from their brew sheets, comprising (i) brew logs, (ii) fermentation logs, and (iii) packaging logs. Designed in an Excel file format by the former head brewer, the preset fields in these logs for each brew were edited digitally before brewing (e.g., recipes), and printed out and clipped together to a board, which brewers carried around the facility to record measurements and comments throughout the brewing process. The key variables for these brew sheets included: the duration of each step as well as the temperature and pH at each step (for brew logs); daily readings of the gravity (a measure of the beer's density of fermentable sugars), pH, and temperature (for fermentation logs); and volume, dissolved oxygen (DO), and CO2 levels (for packaging logs).

The brewers updated the inventory of ingredients, supplies, and packaged beer digitally in a Google Sheets file, which they usually accessed on a shared laptop in their office. Most of the measurements and comments recorded on the paper brew sheets, however, remained undigitized. The paper brew sheets were filed away in plastic binders once a brew was no longer active (i.e., when the finished beer was packaged into bottles or kegs). To use the brew sheet data, the brewers relied primarily on these paper brew sheets. Many paper logs were moldy, stained with wort (the liquid extracted from the malt prior to fermentation), or damaged due to the humid environment inside the brewery. The brewers had conceded that they needed to digitize their data and asked the intern to transfer two years' worth of brew log data (measurements, not comments or annotations) to an Excel file. The data were only partially transferred, however, as the amount of time required for data entry was much larger than they had anticipated; the file was hardly used. The brewery thus had no centralized digital database to engage with data in a systematic manner.

The brewery founder (a former dotcom entrepreneur), the management team, and the brewing team all acknowledged that they needed a better data management system to more sustainably and creatively engage with the data that they had accumulated. Management agreed to explore an initial use case and phase of "digital transformation" of the brewery's operations with us as part of an ongoing research project. The first phase of the project had two goals: (i) identify the production, use, and flow of data at and around the brewery facility and (ii) generate recommendations and project steps regarding how their data could be structured and what kinds of data science tools and technologies (e.g., dashboards, machine learning techniques) could be introduced to enhance production efficiency and augment how the brewers worked with their data.

3.3.3 Data collection

This paper draws on data collected by the first and second authors over the period of one year during the initial phase of the project with the brewery. The authors were introduced to the brewers by the brewery founder as researchers who would study, with their permission, how their brewing process could be "systemized" through digitization, a topic that some of the brewers were skeptical about as they saw their creative input and flexibility as integral to successful brewing. We clarified that our main interest was learning how they actually worked at the facility and what roles data played in their brewing practices in order to help make their work more efficient and meaningful. We explained that the field data gathered at the facility would be reviewed with the brewing team first and could be used in research publications. Once the brewers agreed to participate in our research, we started collecting data. The sources of data included: participant observation, data artifacts, data workshops, and interviews.

3.3.3.1 Participant observation

The first and second authors visited the brewery five times over the span of one year. The duration of each stay varied between several days to two weeks. Our first visits focused on getting to know the brewers and gaining a sense of their routines at the facility. As they became more comfortable with our presence, with their permission, we shadowed the brewers and observed their work practices, asking questions where necessary; had meals and snacks with them; helped them organize their storage room or load their delivery trucks; and participated in their discussions and meetings with their permission. The marketing manager and the distribution manager frequently visited the facility to check the delivery status and discuss with the brewers, both formally and informally, their production schedule and strategies to expand distribution channels. We also sat in on these meetings, audio-recording their conversations with permission. Throughout our fieldwork, we conducted informal interviews with the brewers and the intern, made extensive notes, and photographed and video-recorded their work practices and settings, also with their permission.

3.3.3.2 Data artifacts

A range of data artifacts played an important role in our study. In addition to observing how the brewers engaged with the brew sheets throughout the brewing process, we digitally scanned 1,200+ paper brew sheets stored in the plastic binders. Studying the numbers and annotations on these sheets helped us gain richer insight into the patterns of their brewing practices as well as how they responded or failed to respond to contingencies and mistakes by engaging with the brew sheet data. When we found interesting numbers (e.g., values that were outside of the target range or which were not recorded) and annotations on the brew sheets, we asked the brewers to recount in detail what they remembered about that particular brew and similar cases, if they could think of any. Brew sheets served as an important memory and narrative aid in our formal and informal interviews with the brewers, as their stories were often anchored to the data recorded on these sheets.

We asked the brewers to give us a guided tour of their digital data repositories (e.g., folders on their office laptop and Google Drive), walking us through how the different files (e.g., recipe templates, inventory records) in these repositories were created, edited, and otherwise used. We also were given access to the brewery's ERP system, which was being minimally used by the brewing team, but more extensively by the marketing and management teams, to examine what kinds of data variables were captured, how they were used and communicated across different teams, and how comfortable people at the brewery were with the system.

3.3.3.3 Data workshops

To gain a more systematic understanding of how the brewers used data in their work, we gathered all data artifacts that the different teams used with their permission. We then extracted and compiled all data variables (220+ variables) listed in them, and held three “data workshops” with the brewing team. In our first workshop at the office, we put up brew sheet templates on a projection screen and asked the brewers to walk us through all tasks that they would perform from pre-brewing to packaging stages. For each task, we asked what data they engaged with and how, as well as any comments that they had on that particular engagement, making reference to the brew sheets. The workshop lasted around 60 minutes including a short break, and was audio-recorded.

In the second workshop, we put up the list of all extracted data variables on the projection screen, grouped by the tasks that they corresponded to (e.g., invoicing, mashing, fermentation, packaging). We picked “yield” — the percent of wort that is converted into packaged beer — as a specific topic to guide their descriptions of how they used data variables. We went through the variables on the list one by one, and asked them to identify which variables they thought were potentially relevant to yield and why, and rate the impact of that variable on yield (minor, moderate, or significant impact). When a relevant variable was not being measured or recorded in practice (due to reasons including lack of measuring equipment and safety issues), we asked the brewers whether it could be deduced from other variables on the list or if there were proxies. This workshop lasted around 80 minutes including a 10-minute break, and was also audio-recorded.

In the last workshop, we presented a draft database documentation and database schema to the brewers for their comments and feedback regarding clarity and validity. We also discussed a list of findings on the patterns of their use of data, potential problems in the context of building data science tools, and our recommendations to address these challenges. The discussion lasted around 90 minutes and was both confirmatory (patterns and problems) and exploratory (the feasibility of recommendations). The workshop was audio-recorded with their permission.

3.3.3.4 Interviews

Before we entered our site, we conducted a series of exploratory semi-structured interviews with the brewery's management team (the CEO and the founder) to gain an overview of the brewery's operations and aspirations, and the key data-related problems that they wanted to address. In addition to formal and informal interviews with the brewing team during our fieldwork, we interviewed the marketing manager and the distribution manager in person as well as the former head brewer via WhatsApp, focusing on their experience of using data. Given the potential cultural and organizational particularities of our setting, the second author also interviewed a former head brewer and a facilities manager of a major craft brewery in the US at their facility to understand the types of problems that they faced in their use of data. These semi-structured interviews lasted between 40 minutes and 70 minutes, and were audio-recorded.

3.3.4 Analysis

Our study builds on interviews, videos, audio files, photographs, documents, online chat logs as well as field notes relating to the brewers' work practices and physical settings of TheBrewery. When including photographs in this paper, we blurred out any identifiable logo of TheBrewery. Interviews, videos, and other audio files were transcribed and, where necessary, translated from Korean into English by the first author, who is a native Korean speaker. These transcripts were anonymized and edited lightly for readability. In addition to field notes, we produced notes on the scanned brew sheets, digital files, screenshots of data repositories, and the ERP system as well as photographs of the physical settings of the brewery.

We followed a situational analysis approach (Clarke et al. 2015). The initial set of these transcripts and notes were open-coded by the first and second authors to explore potential themes, which were then discussed with the other authors in project meetings. The coding was iteratively revised. The coding and discussions informed the subsequent field visits and interviews, in which we examined the brewers' engagement with data and associated challenges in further detail.

Given that one of our aims was to provide recommendations for the potential introduction of data science tools to the brewers, we needed to understand how the brewers' use of data was situated in their work practices and settings. In the situational analysis, we analyzed our data to map out and understand the complex web of relationships among key human actors (e.g., the brewers, other teams), non-human actors (e.g., brew sheets, brewing materials, and equipment at and around the brewery), as well as organizational and technological elements (e.g., brewing processes, the dynamics among the brewers, their identity as craft workers, and patterns in their use of paper vs. digital data) that may affect the brewer's data science activities. We then examined how these relationships played out in the brewing cases we could observe in person or reconstruct based on the brew sheet data and the brewers' accounts. In the iterative process of data collection and analysis, we modified our focuses and questions according to the evolving themes, and consulted our tentative themes and findings with the brewers for their comments.

In our analysis, the brewing cases that we observed or reconstructed helped us synthesize the findings from our participant observation, brew sheet analysis, data workshops, and interviews. Focusing on how the brewers' use of data unfolded in these cases enabled us to connect multiple interrelated sets of findings on how the brewers used different types of data before, during, and after brewing, how contingencies affected their use of data, and how the brewers wanted to use data. We selected among these cases and combined similar cases to present three vignettes to capture the range of the brewers' data use at the brewery.

3.4 Results: how craft brewers work with data

In this section, we present three vignettes that surface how the brewers used their data as materials and circumstances in the brewery continually shifted. The first vignette describes using data to improve an existing wheat beer, highlighting problems of unclear interdependencies among variables as well as multiple contingencies and changes during brewing. The second vignette describes using data in developing new seasonal beers, highlighting problems of limited prior data for guiding recipe development and evaluating and improving the beer. The third vignette describes transforming the style of a beer during brewing, highlighting problems of setting new goals and shifting courses of action on the fly. Taken together, our analysis of the vignettes identified processes through which the brewers worked *with* data — situating otherwise abstract data and their relationships in how and where they worked — to deal with changes in the data.

3.4.1 Improving a wheat beer

This vignette illustrates the brewers' efforts to improve the quality of an existing wheat beer by using their brew sheet data. Detecting what was wrong and deciding on how to make improvements to the existing beer was challenging in that there were multiple variables that could have anomalies, and the interdependencies among these variables were unclear. The brewers' ability to modify and evaluate data about a variable was limited by changes in how the data could be interpreted due to contingencies or mistakes that arose during brewing. To improve the beer recipe and modify a brew of the recipe in progress, the brewers used their paper brew sheets as a conceptual device and memory aid that reflected the sequence of their brewing practices, enabling them to make logical connections among variables and recall the circumstances surrounding a particular brew.

3.4.1.1 Before brewing: discussing changes to the existing recipe

One of the brewery's four core offerings was a wheat beer. Once the most popular offering, its sales had been declining over the past few years. With a new batch of wheat beer scheduled to be brewed in a few days, the brewers gathered in the office of the brewery to discuss what needed to be changed to reverse the decline in sales. The discussion was driven by the feeling among the brewers that the decline in sales had something to do with the beer's existing recipe, which had been developed by the original head brewer HB1, who no longer worked there. The brewers raised several possible issues with the existing recipe. One was that the recipe used a type of yeast that was typically used for a different style of beer. However, the brewers remarked that they had not changed the yeast as their wheat beer had been selling fine, and they feared that a change might result in an unexpected taste. Another issue was that the recipe called for chrysanthemum powder but that, even though they were now making a much bigger volume per batch with the change of their equipment, the original amount (300g) had not been increased proportionally. The consensus was that the wheat beer currently had an "overly subtle" chrysanthemum flavor and boosting the chrysanthemum levels might improve the beer's appeal.

With B1 and B4 scheduled to brew the first two (out of three) batches of the wheat beer the next day, they gathered at the office to discuss how to modify the existing recipe. B3 was going to brew the third batch but was cleaning up the facility with B2 and I1. B4 grabbed and flipped through this year's brew sheet binder, and took out the latest wheat beer brew sheets from about one-and-a-half months prior. He spread out the three brew logs on the table. Browsing through the brew logs, and building on their earlier discussion, B1 and B4 first decided to double the amount of chrysanthemum to 600g to see "if anything changes" without making detailed calculations. The previously used hops (called "nugget"), however, were no longer available. They decided to use the same amount of a similar type of hops, called "ella", instead.

B1 and B4 inspected the paper brew logs for a while to see whether to make additional changes. B4 pointed with his finger at the values for "actual final gravity" - the amount of unfermentable sugars in a brew at the end of fermentation — of two of these logs (1.2 and 1.6), which were below their "target final gravity" of 2.5. He commented that finishing with a gravity under 2.0 was an issue. They wanted to taste their previously-brewed wheat beer but there was none left at the brewery. B4 also pointed at the "end runoff pH" values (5.97, 5.94) — the pH of the unfermented brewing liquid after it is initially drained ("run off") from the malt grains — on two of the three logs. The end runoff pH value on the third log was 5.8. The expected value was around 5.8 and a pH value over 6 would lead to a bitter brew. They discussed "topping up" (adding hot water) to lower the pH.

B4 said: "If it starts going to...over 5.8 (pointing at 5.97 and 5.94), you should be topping up...but if this [the pH value on the third log] is the right number, then, we don't need to change anything…Well, let's just brew one more time, just like this, and we'll keep an eye on the final."

"But," commented B1, comparing the three brew logs, "the results could be different because...these (pointing at three "on boil pH" numbers) are different."

"Right…" B4 looked still uncertain of the explanation.

"Let's just brew like this one more time and see what happens," B1 said.

"Alright, cool," B4 said, still examining the brew sheets.

B4 made the two changes agreed on earlier (doubling the chrysanthemum powder to increase its flavor and substituting a similar type of hops due to supply issues) to the Excel recipe file on the laptop. It was the first time that he actually had looked at the laptop screen. In fact, when searching and reflecting on their brew-related data, the brewers almost exclusively engaged with the paper brew sheets rather than their digital versions. Reasons they gave included that the paper sheets allowed them to rebuild “a scenario of what happened with a brew,” “see everything I need all at once” as well as being able to navigate through numbers and make notes or tentative changes without the risk of “messing with files” such as accidentally deleting or entering numbers.

3.4.1.2 During brewing: contingencies arise

Early in the afternoon of brewing the first of three batches for the recipe, B2 noticed that they did not have any free fermenters to store the other two batches, which were scheduled to be brewed over the next two days. The brewers went through the fermenters and bright tanks (tanks that store finished brews prior to packaging), checking the existing batches’ status on their paper brew sheets that were attached to clipboards and hung on the tanks. They decided on a plan to transfer one batch stored in a bigger bright tank to a smaller one that day, and then to transfer another batch, which was almost done fermenting, from the fermenter to a small bright tank during the first shift the next morning. The transfers would free up a fermenter and a big bright tank for the new batches.

This plan, however, required them to use a lot of hot water for cleaning and disinfecting the fermenters and bright tanks, which could interfere with the hot water needed for brewing the third batch of the wheat beer. The schedule would be tight, but the problem seemed to have been solved. On the double-brew day, B1 was going to brew the first batch according to the updated recipe. He cleaned the bright tanks with hot water. When B1 was mostly done, B1 and B4, the second brewer, sat down together with B1’s paper brew log to discuss what needed to be adjusted based on the result of the first batch. The mash pH of the first batch was 5, which was below the target range (5.2-5.3). B1 wrote in the notes section of the paper brew log that the grain transfer speed was also faster than usual, and B4 took note of that.

After lunch, B4 started "mashing in," a process in which the malt for the beer is mixed with hot water to begin to extract fermentable sugars from the malt. Soon, he started yelling: "Who's using hot water?!" There was no hot water left, and as mashing had already started, there was no turning back. He was upset with the other brewers, saying that they were careless with the use of hot water. B1 and B2, who cleaned the tanks that morning, tried to calm him down and solve the issue. Knowing that the cold water that flowed through the hose connected to the heat exchanger came out heated and went into the water tank, B1 and B2 slowed the water speed to make the water more heated so that its temperature could be closer to the target temperature of 70°C. The actual temperature of the water ended up being around 60°C. B4 managed to finish brewing the second batch, using the lower temperature water.

3.4.1.3 Handling contingencies during brewing by making ballparked modifications

The third brewer involved in the wheat beer was B3. Before his shift began to continue the brew with the third batch, B3 and B4 together looked at the numbers on the two paper brew logs from the previous two batches (B1's and B4's). Because of the use of the lower-temperature water, they worried that it would produce more fermentable sugar than usual, which would change the rate of fermentation. They decided to increase the amount of the "caramel pils" malt, a type of malt that contained non-fermentable sugar, from 25kg to 75kg, while reducing the amount of "pilsner" malt (200kg → 175kg) and wheat (150kg → 125kg). They also doubled the amount of the "ella" hops from 200g to 400g. This new ratio, they explained, was calculated based on their knowledge of brewing science principles, and their experience with the recipe. They would be able to evaluate the result once the three batches had been combined and fermented. In about two weeks, the combined batch completed its fermentation, with a final gravity of 1.2, which still was much lower than their target of 2.5. The brewers said that the final wheat beer tasted “fine” nonetheless, and that they might want to try the updated recipe again for the next brew, hopefully without unexpected changes, to “really see” the effects of changes

Analysis: Unclear interdependencies among variables. The waning sales figures for the wheat beer prompted the brewers to modify the existing recipe. Yet the brewers were cautious about modifying variables since they did not want to create inconsistencies in the beer. Determining what variables needed to be changed and by how much was challenging. Over the years, the materials and brewery settings had changed (e.g., the departure of the head brewer who created the recipe, the bigger batch volume (thereby reducing the ratio of chrysanthemum), the yeast type). In analyzing the prior data on the same wheat beer recipe, it was not clear which variables would be affected. The brewers detected anomalies by manually comparing a small number of measurements across different batches or against target values. Yet, as seen in the discussion between B1 and B4 on anomalies in the data for the “final gravity” and “end runoff pH” variables, the brewers were unable to trace the source of these anomalies, or make systematic connections between them or with other variables. The brewers decided to modify a variable that seemed important but also more easily controllable — boosting the flavor of chrysanthemum by increasing its amount rather than changing the yeast type or the “top-up volume” for which the results would be less predictable.

Analysis: Multiple contingencies and changes. Other than changing the amount of chrysanthemum powder and substituting a similar type of hops, the brewers initially planned to keep the recipe the same and monitor some of the suspect variables (“let’s just brew one more time like this and see what happens”). When it came time to actually brew, multiple contingencies and changes arose: the grain transfer speed was faster than usual; an initial lack of hot water during brewing the second batch led to a lower-temperature water temperature; and the malt bill for the third batch was changed to account for this contingency. These multiple unexpected contingencies and changes during brewing meant that the data on these batches were no longer readily comparable with the data from other batches of the same recipe.

Analysis: Situating data in practice with paper brew sheets. The vignette illustrates how the brew sheets themselves served as a type of conceptual device that logically bound the relevant variables according to how the brewers made sense of their brewing practices (e.g., starting with the recipe, measuring different variables during the sequential brewing process of mash, sparge, boil, cool, etc.). A further observation is that, while their digital brew sheets had the same content and format as their paper brew sheets, the latter more easily travelled around the physical settings of the brewery by being placed on clipboards, thereby retaining the context of the batches to which they were literally attached (e.g., to the mash tun during the mashing process, or to the fermenter during fermentation). Their memories of the physical movements of the ingredients, water, batches, brew sheets, and the brewers themselves across the brewery throughout the brewing process helped them rebuild a scenario around a brew with otherwise abstract and discrete data points. Interpreting their brewing data to modify the wheat beer recipe both in advance of and during brewing was thus intimately tied to the brewers’ ability to situate the data within their work practices and settings.

3.4.2 Developing seasonal beers

The second vignette illustrates how the brewers developed and improved new recipes for seasonal beers, where prior data and experience were inherently limited. The brewers felt that the data variables and target values of the recipes for these "imagined" beers were abstract. Given the lack of prior data and experience, the brewers situated the abstract quantitative data by drawing on their knowledge of operations specific to the brewery settings, conducting on-the-spot sensory evaluation (e.g., how the brew looks after mashing), and retrieving data from brews of analogous recipes. The vignette was reconstructed from the brew sheet records and the brewers' accounts.

3.4.2.1 Developing the recipe for an "imagined" beer

The brewers experimented with recipes using local, seasonal ingredients to develop specialty beers to complement their four core offerings. As one example, the city where the brewery was located held an annual festival in early summer where changpo - a root used in traditional Korean medicine - was an important part of the celebration. The brewers decided to develop a saison-style beer made with the changpo root that would be served at the festival and their brew pub during the festival period. As another example, the brewery developed a rice lager as a summer seasonal, driven by the local government's initiative to promote the rice produced in the region. To get funded under the initiative, local rice needed to account for at least 40% of the total grain bill of the beer. B2 commented that, since these seasonal recipes had to be developed for the first time, a challenge was that "all the characteristics are very abstract. We have this imagined beer, imagined flavor in our heads."

The brewers drew on their brewing experience and knowledge of the local ingredients and potential customers in discussing what beer style would best complement the defining ingredients, such as "saison" for the changpo beer and "lager" for the beer made with local rice. Recounting how they decided on the beer style for using the changpo root, B4 explained: "Changpo is earthy and it's got these mildly medicinal characteristics to it. Would we put that in a hazy IPA? No. Mango juice and medicinal herbs do not go together. Saison is known to be spicy and herbally, and has all these characteristics that would agree with changpo." B1 commented on the rice lager, another summer seasonal: "It's hot...so you'd want something light and drinkable, and lager would be good for that. And this helps us determine our ABV (alcohol by volume) and the strength of the flavor."

When the intended style, ABV, and flavor were determined, the first go-to reference in developing the recipe were the data on their previous brew sheets that had specifications similar to what they had in mind. B1 explained. "If I want to make [a seasonal] 4.8% ABV, I would try to find the brew log for a brew whose ABV was, for example, 4.6 and check the recipe...and try to adjust it little by little to get what (ABV) I want...because I have the experience (with that brew) I would know why we got the numbers (on the brew log)." They then decided on a list of grains and calculated their ratio and targets that would help them produce "a beer that is close enough to what [they] imagined," using formulas compiled in an Excel file by HB1 and existing software programs for designing beer recipes.

Similar to deciding on a style, the brewers also discussed what varieties of hops would agree with the beers. They ballparked the hop amounts based on their experience. When they used software programs for calculating the numbers, their interpretation of these numbers was situated in their operational knowledge of the brewery settings (e.g., "100kg of malt roughly equals to X% ABV in principle, but in our previous brews, the number tends to be Y given the operational efficiency of our brewing process"). Another factor that played out in developing the recipe was the brewers' knowledge of their inventory and production schedules. The brewers commented, for example, that different ingredients came in differently-sized bags (e.g., 20kg or 25kg) and that they tried to use the whole bag or a half bag depending on which beer was scheduled to be brewed next so that they could keep ingredients fresh for each brew and manage inventory effectively. B2 commented on why they used a range of data and knowledge: “After all, we need to develop something new while minimizing risk, so we bring in what we have to concretize this new beer as much as we can.”

3.4.2.2 Trying out a recipe and seeing what happens

Despite developing seasonal recipes by drawing on prior data, the brewers questioned the quality and value of these data. First, dealing with new or infrequently-brewed styles and ingredients meant that they did not have enough prior data to determine specific targets for the brewing and fermentation process variables or precisely predict the outcome variables for evaluating the final product. “You have, say, ripe persimmons, and you can’t really predict how their color or flavor will materialize after fermentation,” B2 explained regarding a prior seasonal persimmon ale recipe. “You don’t really know how much of them you would need to make [the brew] look or taste like a persimmon beer or at which point you need to put them into the batch...not to mention, this year’s persimmons could be different from last year’s and different producers might process them differently.”

In addition to limited prior data, the data that they did generate could be unreliable. The production volume tended to be smaller for seasonals (one or two batches), so the brewing process often involved a single brewer, rather than the three brewers typically used for brewing core offerings that used the three-batches-in-one-brew system. Comparing seasonals to core offerings, B1 explained: “The problem (with the data on seasonals) is that there’s only a few data points and the variance tends to be larger, but we don’t know what caused the variance...we don’t really know what to get out of it.” A seasonal beer would be brewed once a year such that the brewers had trouble recalling what happened with the previous brew in exact detail. The brewers in fact gave us slightly different stories about, for example, whether certain ingredients were changed for a particular batch. These uncertainties and unreliabilities surrounding the limited prior data on seasonals led the brewers to rely also on their senses to evaluate the process. “We don’t really have guidance on the process, I trust what I see in the actual batch more than abstract numbers,” B2 said. “I physically check the batch along the process, like...its thickness, color, or taste, and see if the measurements make sense.”

3.4.2.3 Evaluating and improving the imagined beer

Given that the seasonal beers were brewed from new or rarely-used recipes, what the resulting brew was supposed to taste and look like was also not well-established. The brewers’ evaluations of the seasonal beers depended in part on imagined rather than well-defined and measurable outcome variables. The criteria that the brewers drew on ranged from sales figures, customers’ comments, the suitability for a given season, uniqueness, the complexity of the recipe, and cost- or time-efficiency to “something that everyone around says is a really good beer.” In the cases of the changpo saison and the rice lager, the results differed from what the brewers had imagined, but were nonetheless viewed as “good enough to sell” as a seasonal.

Even if a seasonal beer was considered overall to be a success, the brewers tried to make changes to the recipe to improve its quality. Lack of data, however, constrained their ability to make changes efficiently and systematically. For example, when B4 brewed the changpo saison, he first asked the brewers who had brewed the beer the year before about what they remembered. The amount of changpo powder used in the original recipe was 300g and the medicinal flavor was “too much.” B4 wanted to make it more subtle and cut the amount to 250g based on his hunch. “If we had a better log, I could’ve ballparked it better,” he commented. “In hindsight, we could’ve experimented more systematically but especially on that small scale, you just wish for something that would be drinkable, we gotta sell this.”

Similarly in the case of the summer rice lager, it sold well the previous year and its recipe was relatively simple, so while no longer funded by the government, the brewers had decided to offer the seasonal again. The brewers felt that the barley flavor in the previous year's brew was "too subtle," but as the beer was marketable, they did not want to make a dramatic change. They decided to reduce the amount of rice from 40% to 36%, a decision that B1 and B2 made based on "personal experience and knowhow" as well as their knowledge about their current inventory levels, rather than systematically analyzing and experimenting with the malt-rice ratio. The brewers reported that the result was successful but could still have been better.

Analysis: Lack of guidance on processes. In "imagining" new beer recipes, the brewers were able to use their experience and knowledge of brewing science to select the beer styles and specify targets for core variables such as ABV. They were then able to use these styles and targets to identify brew sheets from prior batches of analogous recipes. They used the data in these brew sheets to calculate specifications for the amount and type of malts and other ingredients, as well as target values for process variables for the imagined beer recipe. Yet even when a target for a process variable was considered out of range, the brewers often did not have a clear idea of what kind of corrective action could be taken or how that would affect the outcome variables of the end product. The limited prior data and experience highlighted the brewers' desire to get guidance about processes from their data. Not only were data from previous brew logs of analogous beer recipes scarce and high in variance, but the brewers had difficulty in extrapolating precise insights from the data that did exist.

Analysis: Which lessons? When brewing an "imagined" beer, the brewers found it difficult to know in advance how the end product was supposed to look or taste, or how to evaluate the end product, even with the predetermined targets. What was seen as a "success" was loosely defined based on a wide range of factors, and it was rare that a seasonal was considered unsellable. The small amount of data not just regarding the process, but also regarding outcomes of the beer recipes led to the need for subjective judgment from the brewers themselves. The problem was not so much in evaluating the result but in interpreting and documenting the data from that experience in a way that could potentially be used for identifying outcome variables for improving the seasonal beer the next year.

Analysis: Situating abstract quantitative data in the brewery settings. To develop a new recipe, the brewers tried to define an imagined beer (e.g., a refreshing, drinkable rice lager for summer, or a light changpo saison for an annual festival) in terms of measurable inputs (a list of grains and hops and their ratios) and outputs (target gravity, ABV, and pH). In defining these specifications and targets, they drew on a range of resources including: prior data from brew sheets, recipe programs (for calculating ratios and targets), knowledge of the qualities of the ingredients they used, beer styles, their inventory, the facility's operational efficiency, as well as sensory evaluation (for monitoring the thickness, color, and taste of a batch in process). The brewers' use of these resources highlighted their tendency of situating abstract quantitative data in the particular settings of the brewery. For example, choosing the grain ratio was dependent on the context of how the ingredients on hand were packaged and how they could be kept fresh given the current production schedule. Similarly, when the brewers had set a target of 4.6% ABV for the changpo saison, they referred to a previous brew log to see what brewing a 4.6% beer involved in terms of not only the recipe but how the brewing process unfolded in the physical settings of that brewery and what the recorded numbers meant.

3.4.3 Transforming a beer

In this vignette, we examine how the brewers used data to transform the style of a brew in the middle of the brewing process and without well-defined targets in order to salvage a batch. While transforming the brew eventually was a success, the brewers found it challenging to find effective corrective actions after the initial detection of anomalies in the data. The vignette was reconstructed from the brew sheet records and the brewers' accounts.

3.4.3.1 Transforming styles of beer during brewing

About a year prior, B4 had been brewing a batch for a double IPA recipe (IPA, or "India pale ale" is a style of beer popular in craft brewing that is characterized by a hoppy flavor and relatively high alcohol percentage; a double IPA is an especially high alcohol version of the style). Early in the brewing process, B4 recorded data for the batch's "start runoff gravity," a variable that refers to the fermentable sugars in the brewing liquid when the liquid is first extracted from the malt. The target start runoff gravity for the double IPA recipe was around 28.0, but B4's measurement was 14.7, way out of range. Puzzled, B4 went to HB1, the head brewer at the time, and said that the number didn't make any sense. B4 and HB1 went up to the brew deck (a small platform up a set of metal steps that offered an overhead view into the brewing containers). They identified a major problem with the malt mill auger, the conveyor system that transports grains from the mill to the mash tun (the large stainless-steel container where the brewing liquid was initially heated). They could not halt a brewing process that had already been started, nor could they afford losing the whole batch. Given the low gravity measurement, they agreed that the batch might be salvaged as a session IPA (a lower-alcohol version of an IPA). Session IPAs were not part of their core offerings, and they did not have a recipe with targets specific to their brewery. They made changes to the hops to align with a typical session IPA recipe. He left a note on the brew log: *"Problem with auger. Overshot volume 130L. Silo stuck up with about -30% malt so gravity is much lower. Adjusted recipe accordingly - will likely have a session IPA or similar.*

3.4.3.2 The "SHIT SHOW"

After the batch was transferred to the fermentation tank, B4 began the fermentation log by writing "SHIT SHOW" in the day-zero comments section of the log. "I was like, 'no, I don't know if this is going to work'," he said. The brewers were not sure how to proceed with dry hopping — a process in which hops are added to the brew after the initial phase of fermentation — to transform the brew into a reasonable approximation of a session IPA in their existing settings. Without specific guidance or targets to follow, they ended up hopping the beer as they would hop their double IPA, given that they were more familiar with this process, even though their session IPA was a 4% beer and their double IPA was a 9% beer. B1 and B2 took turns to dry hop the brew twice, and HB1 and B1 recirculated the brew with a pump for six hours to make it "closer to a session IPA." The result was "horrible, horrible beer," B2 and B3 recounted. After fermentation, the brew was left in a storage tank with no immediate plan for packaging.

3.4.3.3 Re-transforming the beer

About two months later, the brewery happened to have some leftover sweet cherry and guava purees from making a seasonal beer. On a hunch, HB1 convinced the team to add the purees to the "horrible" session IPA, hoping that it might make it drinkable. It didn't work — the beer still tasted horrible. As a final effort, the brewers recalled having brewed an apricot sour beer recipe as a seasonal beer and decided to try to turn the brew from a session IPA into a sour beer by referencing that recipe. To do so, B2 and B4 brewed a half apricot sour beer batch (1,000L) and blended it with the so-called "session IPA" with the cherry and guava purees. The result was a 4.1%

beer with a final gravity of 1.7 — a session IPA's target final gravity is usually around 1. "Surprisingly, it tasted good. It became our 'super sour fruited session IPA'," B4 commented. The accidental seasonal offering was a moderate success. Once the summer was over, however, the sour beer sales went down, and they still had several leftover kegs at their facility. Although the brew was more than a year old, the beer had not been fully shipped and thus was considered "active." Its packaging log was still on one of the kegs, not in the binder, making the data on it practically invisible to the brewers. While the data on the brew sheets served as a memory aid, much of what happened in the process was also not recorded on the brew sheets themselves.

Analysis: Moving targets. In this vignette, a considerable portion of the grain was lost due to the equipment failing to work as expected, which meant that some of the variables for the brew also deviated from their target values in the intended recipe. To salvage the brew, the brewers had to come up with modifications beyond adjusting amounts of malt or hops, transforming the style of a 2,000-liter brew in the middle of the brewing process. The brewers drew on their knowledge of different beer styles and their properties (e.g., ABV, gravity, and flavor) to find a style that this brew could be turned into, given the measurements that they had so far. They decided that switching its style from a double IPA to a session IPA might be feasible, as the latter was a lower-alcohol IPA that could accommodate a lower amount of malt. The brewers, however, were not experienced in brewing a session IPA and did not have a tried-and-tested recipe that they could confidently use in their material settings. As they continued to change the style of their existing brew (from a session IPA and a fruited session IPA to a "super sour fruited session IPA"), it became difficult to calculate new targets to make the beer taste "right" or "good enough" according to the new style. When the failed session IPA was mixed with a sour beer and purees, the brewers did not have a clear idea of how the re-transformed beer was supposed to look or taste, or how to interpret the data on the results.

Analysis: Shifting processes midway. In addition to the moving targets, the brewers were not sure what or how many of the variables of the brewing and fermentation processes actually needed to be changed in order to achieve an intended transformation successfully. The initial batch specifications mostly reflected those of the initial double IPA recipe, so it was plausible that at least some of the targets for processes should reflect the initial recipe rather than the new style of a session IPA - after all, they could not start the brewing process anew and mostly had to work with the existing brew. When the beer's initial transformation into a session IPA failed, the brewers had to be flexible and resourceful, as their use of data could not lead to reliable predictions of the outcomes of possible corrective actions.

Analysis: Making informed dice rolls. The challenges of using data to transform the style of the brew led the brewers to err on the side of caution in making modifications to process variables (e.g., sticking to the same amount of hops as in the double IPA recipe). The brewers were able to get beyond their initial failures with the session IPA by instead making informed dice rolls that eventually transformed the brew into an unexpectedly sellable sour beer. These dice rolls were experiments that arose from hunches (e.g., making a session IPA), taking advantage of materials on hand (e.g., adding the cherry and guava purees), and drawing on their brewing experience (e.g., making an apricot sour recipe that they were familiar with). Interestingly, the transformed brew tasted "surprisingly good" and sold well, even though its gravity reading — a key target variable — was off target (1.7 vs. 1.0). Many of the modifications that the brewers made, however, were not recorded in the brew sheets, which limited the brewers' ability to learn from their "dice rolls" to inform their future brews.

3.5 Discussion

3.5.1 Situating data and data science in the practices and settings of craftwork

This paper builds on emerging research in CSCW and HCI on data science tools to support domain experts (Gautrais et al. 2021; Gil et al. 2019) by studying their use of data through the lens of craftwork. The three vignettes illustrated how a team of craft brewers addressed problems of changes in their data when improving an existing recipe, developing a new recipe, and transforming the style of a brew. Our analysis of these vignettes extends recent insights into the situated interpretations and manipulations of data in data science (e.g., Passi and Jackson 2017). In particular, our analysis identified processes by which the craft brewers worked *with* data — how they were able to grasp and control complex and unpredictable materials and brewing processes by situating changes in the values, target ranges, and expected relationships of their data *within* the material practices and physical settings of their craftwork.

Craft is often considered irrelevant to data science, where much of the focus has been on the statistical analysis of big data sets in scientific research, large technology firms, or industrial settings. Recent calls for expanding access to data science tools (Gautrais et al. 2021), however, point to the ubiquity of data and its use in humans' domain-specific work practices and settings (e.g., Boukhelifa et al. 2017). Craftwork is especially salient to highlighting issues in expanding access to data science since craftwork is less regimented and industrially-driven, and thus less prone to the kind of controls necessary to an industrial or techno-industrial understanding of data science that may overly abstract data from its situated use.

Our findings offer theoretical and practical contributions for studying data science through the lens of craftwork. First, we discuss how a focus on craftworkers as a type of domain expert is distinct from theories of the work of technical data scientists in terms of the view of data that is assumed (situated vs. abstract), the specific purposes for engaging with data (guiding processes over predicting outcomes), and the overall goals of using data (flexible control vs. increasing precision). Second, we discuss how a focus on working *with* data points to distinct challenges for the design of data science tools. We conclude with the limitations of our study and future directions. By highlighting how and why data science for craftwork can be theorized and designed for, our study contributes towards integrating domain expert-driven understandings into the emerging research on data science in CSCW and HCI (e.g., Banovic et al. 2019; Gil et al. 2019).

3.5.2 Theorizing how domain experts work with data

3.5.2.1 Situated vs. abstract data

Much research on data science in CSCW highlights practices of developing and interpreting statistical analytics tools such as machine learning models or algorithms that operate on data, but gives less attention to the practices of engaging and thinking with the data themselves. This focus on analytic models or algorithms is consistent with the priorities of technical data science and machine learning practitioners on developing generalizable machine learning models that learn to make accurate predictions "independent of which dataset is used" (Hohman et al. 2020, p.1; Sambasivan et al. 2021]. Changes in data are seen as detrimental to model performance and need to be tightly controlled (Gama et al. 2004; Sculley et al. 2015; Williamson and Henderson 2021). The data are rendered abstract and static in that a single data set can be statistically manipulated and modeled in many different ways. This emphasis on abstract data, as Ribes et al. (2019) indicate, elides a central tension in data science between the priorities of technical data science on developing models that are independent of domains and the socially and materially situated practices of a specific domain (Passi and Jackson 2017; Suchman 1983; Lave 1984).

In our study, the craft brewers searched for a logical structure of how variables could be put together according to their work practices and settings, with which they could recall and reconstruct a brewing scenario. Measurements of these variables (e.g., grain supplies, equipment, operations, or even weather conditions), while subject to a statistical distribution, were specific to each batch, and often not easily comparable across different brews. Extracting abstract insights from the data would have been inadequate for TheBrewery, as a specific site of practice, to support domain experts' use of the data. The brewers wanted tools to work *with* data — to integrate data as part of their material work practices and physical settings — to leverage their knowledge and experience of brewing and of the particular brewery.

As one example, the brewers often preferred to work with data on paper brew sheets rather than the same data in digital files. The paper version physically traveled with a brew in the brewery as well as provided a visualization of the structure of variables (consistent with e.g., McCullough 1998; Feinberg 2017) that reflected the order of their brewing process (i.e., preparing grains, mashing, sparging, boiling, cooling). By being situated in the material practices and settings, the paper brew sheets provided a tangible resource to explore and reason with multiple variables and their complex relationships to figure out what was happening and why across multiple brewing stages.

The "interdependence" between paper brew sheets and the brewing process points to the challenge of designing digital artifacts for data science that work better than or complement paper sheets (e.g., Sellen and Harper 2003; also Heath and Luff 1996; Dolata and Schwabe 2017; Marathe and Chandra 2020). Data science for working *with* situated versus abstract data builds on a view that the material context leads to variations in the meaning and relevance of data, and these variations need to be designed for rather than controlled (e.g., Amershi et al. 2014; Hohman et al. 2020).

3.5.2.2 Guiding processes vs. accurately predicting outcomes

Data science is commonly understood in terms of making statistical predictions about outcome variables from large data sets that are then translated into actionable insights (Dhar 2013). Desire for greater accuracy of these predictions incentivizes a search for simpler models that are able to optimize the prediction metrics by taking into account a small number of variables (Dhar 2011; Meinshausen 2007; Forster 2002).

In our study, the targets for outcome variables were often abstract and difficult for the brewers to engage with; in some instances, the end product could be considered a success even when output variables were well out of their target range. To get to actionable insights, the brewers needed to make sense of relationships across a large number of variables (e.g., water temperature and final gravity, the amount of chrysanthemum powder and the flavor after fermentation). As a result, the brewers were less interested in optimizing the accuracy of predictions regarding any particular output variable. Working with data instead involved getting a sense of multiple patterns in the brewing process as it unfolded in its setting.

Getting the process right matters in craftwork in that such work depends on a continual “dialogue” between craft workers and their materials, or more specifically between craft workers’ images of the materials and properties of the materials (Sennett 2008; Marchand 2016; Keller and Keller 1993). This dialogue, according to Sennett (2008, p.272-3), requires consistent rules for engaging with unpredictable materials and increasing complexity in rule making through repetition and modification. Our findings show how this continual “dialogue” extended to how the craft brewers worked with their data. The brewers analyzed data robustly by using logical relationships, and not just statistical algorithms, among variables as guidance for finding potential problems and constructing actions for addressing these problems.

3.5.2.3 Goals of flexible control vs. precision

The craft brewers' emphasis on getting the process right based on analysis of situated data led to overall goals of gaining flexible control over materials and practices, rather than achieving precise target values that assumed a data set from a single statistical distribution. The brewers saw variance in data as a natural part of their work practices. As contingencies arose in the grain-to-glass process, the materials and circumstances surrounding each brew varied, which made it difficult or impractical to directly compare two brews of the same recipe. The brews hardly hit their target numbers exactly, but the brewers usually considered them "fine" or "good enough" to sell.

Given the variance in measurements and the loosely-defined evaluation criteria, the brewers sought to learn with target *ranges* of values that could serve as guidance for evaluating whether a measured value was expected (within range) or potentially problematic (out of range). The checks on ranges were an important part of keeping their brewing process on track. The brewers frequently ballparked when calculating recipes and target numbers by drawing on brewing science principles and prior experience. They questioned how precise or relevant calculations with software programs could be in the face of continual changes in their material practices and settings, as well as the lack of prior data that could be used for systematic experiments. The brewers' use of ranges or ballparking could be seen as a means to get a sense of flexible control over their materials and practices by leveraging a large set of logical relationships among variables rather than making precise predictions.

Flexible control over resources and work practices is a recurring theme in research on domain experts' work practices in CSCW (Suchman 1983; Luff et al. 2000; Blomberg and Karasti 2013). In craftwork, flexible control over materials is part and parcel to craft workers' identity (Marchand 2016). As Becker (1978, p.865) explains, a skilled craftsperson is someone who "has great control over the craft's materials, can do anything with them, can work with speed and agility, can do with ease things that ordinary, less expert craftsmen find difficult or impossible." Likewise, we propose that, at the heart of a domain expert-driven understanding of data science is the overall goal of giving domain experts a sense of flexible control over their work practices and material settings through their engagement with data.

3.5.3 Designing data science tools to support working with data

In this section, we explore directions for designing data science tools to support domain experts. To do so, we generalize our findings into four interrelated tasks that the brewers performed in working with data: (i) monitoring variables; (ii) hypothesizing relationships among variables; (iii) calculating modifications and new targets; and (iv) evaluating hypotheses and data.

3.5.3.1 Monitoring variables

Monitoring variables involved tasks such as: (i) checking if the measured variables fell within their target ranges (e.g., runoff gravity over 2.0, runoff pH around 5.8); (ii) comparing target values (e.g., expected final gravity) and actual values (e.g., actual final gravity); and (iii) monitoring trends in variables (e.g., in the fermentation process, temperature should be consistent, and gravity and pH should gradually decrease over time). These tasks were mainly for detecting anomalies (e.g., a measurement out of range or a notable difference between a target and an actual) in variables, both during and after brewing. Done manually with the paper brew sheets, these activities were largely limited to one-on-one comparisons between 2-3 brew sheets, as highlighted in the first vignette. The brewers saw these comparing and monitoring tasks as critical to quality control.

In relation to data science activities, a key challenge was deciding on the criteria for an anomaly given that the variables' target values would be a little different for each batch. *Post-hoc* manual comparison made it difficult to analyze multiple variables across multiple batches and detect subtle, irregular, or long-term trends, limiting the brewers' ability to take remedial actions quickly and efficiently. Data science tools may provide support for, for example, real-time pattern visualization and alerts for out-of-range data.

3.5.3.2 Hypothesizing relationships among variables

Once an anomaly or problem was detected, the brewers performed tasks of reasoning about causal relationships among variables that might explain the out-of-range data (e.g., relating an off-target ABV for a batch to the ratio of malt to water) and help generate corrective actions. These tasks were usually performed in the brewery office after brewing by referring to paper brew logs, especially areas within the brew log that grouped temporally related variables together.

A key task was identifying what variables contributed to the outcomes out of their expected range. The brewers did not have an established means of systematically exploring the large number of variables (220+) on the paper brew logs alone. Their use of data for hypothesizing was usually limited to speculating about correlations and their cause based on manually detected anomalies by comparing values of a small number of variables in the most recent two or three batches of the same beer type (e.g., gravity "a little high(er)" and runoff time "quick(er)" than the values in the previous wheat beer brew sheets). Further, whether in use or not, they lacked resources to quantify hypothesized relationships, making them vague and unactionable. This challenge points to the need to support the use of data science techniques (e.g., machine learning) for pattern discovery across batches by integrating the brewers' knowledge and experience in prioritizing and selecting a smaller set of variables ("feature selection") to include in a hypothesis. The implication is that, even assuming an effective machine learning tool for anomaly detection, such a tool would be inadequate for determining which among many interdependent actions to take to adjust the batch.

3.5.3.3 Calculating modifications and new targets

The brewers used their knowledge and hypotheses about the relationships between variables to calculate or ballpark modifications to existing recipes as well as targets for new recipes. These activities were aimed at setting specifications and target values to guide future corrective actions, improving the consistency and quality of beers against contingencies, or creating new beer styles. Before brewing a particular beer, the brewers reflected on the last time the same beer (or similar one) had been brewed, drawing on the data on the brew sheets, the tasting of the beer, and variables that the brewers remembered to be relevant to pay attention to in that brewing process. They also used brewing recipe software programs and supplier-provided grain specifications to calculate ratios for a modified or new recipe. Calculating modifications also happened during a brewing process.

Given the lack of comparable prior data and their habit of using ballparks, key challenges in these modification tasks concerned making reliable calculations and conducting systematic experiments. Modifications were made mostly *post hoc*, limiting the brewers' ability to take corrective action during a brewing process. At the same time, they worried that modifying the recipes might negatively affect the consistency of their beers. The brewers did not feel confident to flexibly manipulate the templates of recipe variables and embedded formulae in the Excel sheets designed by their former head brewer. These challenges resulted in a tendency to keep changes to a minimum, such as brewing things the same way again to “see what happens.” To address these challenges, data science tools may provide automated recommendations for possible specifications or tasks by integrating a large proportion of data on ingredients and their ratios given a beer style and an intended alcohol percentage, integrating data on current inventory levels, grain specifications, and past brew sheet data as well as data from external online brewing data repositories.

3.5.3.4 Evaluating hypotheses and data

Once brewing was finished, the brewers sought to evaluate the effects of changes made to the recipe based on their hypothesized relationships among variables, as well as to evaluate the overall quality and popularity of the beer. The purposes of these evaluations included improving quality control and deciding whether and how to adjust future brews of the same recipe. In the case of seasonal beers that had abstract targets and recipes that were not tried and tested, evaluating their success drew not simply on how much actual measurements deviated from target numbers, but also on sensory evaluation, sales figures, customers’ comments, knowledge on market trends and local ingredients as well as operational efficiency (how easy or cost/time-efficient the brewing process was).

In these data science activities, accurately evaluating the impact of individual variables on the outcome was difficult or impossible due to the interdependence of variables as well as changes made to respond to contingencies. Logically linking different types of data was also challenging (e.g., linking brewing or fermentation process targets to sensory evaluation to sales figures), making evaluations often unsystematic or inconsistent. Data science tools may be designed to support customizable views for exploring relationships in various subsets of variables across multiple brews as well as the integration and visualization (e.g., dashboard) of data across brewing, sensory evaluation, and sales figures.

3.5.4 Limitations and future work

While our ethnographic study of a craft brewery in Korea provided rich accounts of how a small group of craft brewers worked with data situated in their work practices and settings, the small size and cultural and organizational particularities of our site might limit the generalizability of our findings to other settings. We conducted an interview with a major craft brewery in the US to examine commonalities and differences between the two breweries, which confirmed that the types of problems for which the brewers at our site worked with data also existed in larger sites within the craft brewing industry. While our study focused on a particular type of domain experts (i.e., craft workers), we believe that our high-level framework of working with data is applicable to other domain experts' work practices and settings. Our study presents opportunities for future research on how our framework holds across different sites.

Another limitation is that, while this study explored possibilities for domain experts' use of data science tools as part of an ongoing research project, it did not touch on the actual development or implementation of such tools (e.g., machine learning models). In order to build data science tools that could be useful and usable for domain experts [e.g., 31, 30], we believe it is critical to understand how domain experts work with data in their work practices and settings. Future studies may seek to draw on our findings to design, implement, and test data science tools for different types of domain experts, including domain experts who might not be characterized as craft workers.

3.6 Conclusion

This chapter explored a domain expert-driven understanding of data science through the lens of craftwork, analyzing how a team of craft brewers used data as the materials and situations in their brewery continually shifted. The vignettes illustrated how the brewers worked *with* data by situating changes in the values, target ranges, and relationships of variables within their practices and settings. By working *with* data, the brewers were able to gain a sense of control over complex and unpredictable materials and brewing processes. The findings contribute theoretical insight into how domain experts' use of data is distinct from the use of data by technical data scientists. The study also contributes practical implications for CSCW and HCI research on designing tools that support humans' engagement in data science activities in their domain-specific work practices and settings.

As pointed out in 3.5.4, the purpose of this study was to frame domain experts as potential end-users of data science tools and to develop insights into how they work with data. In the next chapter, I seek to connect domain experts' data practices to data science activities by illustrating the course of a pilot ML project conducted at the same site.

4 Understanding Domain Experts' Data Practices for Sustainable Data Science

In this chapter[2], I continue with the ethnographic study at the craft brewery. By analyzing a pilot ML project, I draw a connection between the data practices of the brewers and the ability to sustain data science activities over time at the brewery.

Research on data science has largely viewed data as an abstract input in service of algorithms developed by data scientists. In this view, data science activities are made sustainable by the efficient flow of data to improve the algorithms. Recent studies in CSCW and HCI, however, point to how the effectiveness of algorithms critically depends on sustainably collecting reliable, complete data situated in domain experts' practices and settings. Drawing on ethnographic fieldwork and a pilot ML project at the same craft brewery, this chapter describes three types of situations where brewers' data practices led to unreliable, incomplete data, and how such data practices limited the effectiveness of data science activities. I analyze sources of misalignment between their data practices and data science activities, which are used to offer design implications for sustainability. Extending research on end-user software development that views sustainability as driven by domain experts as *owners of problems*, the study proposes data science research can be driven by domain experts as *owners and users of data*.

[2] This chapter (Jung et al. 2022c) has been submitted to Computer Supported Cooperative Work (CSCW): The Journal of Collaborative Computing and Work Practices and is under review. I proposed the overall framing of the study and led the ethnographic fieldwork. The pilot ML project was led by the co-authors (Tom Steinberger and Chaehan So). I analyzed and wrote the findings from the ethnographic fieldwork, with Tom Steinberger analyzing and writing the findings from the ML project. I integrated the findings and wrote the first draft of the manuscript with feedback from Tom Steinberger.

4.1 Introduction

Data critically shape how data science and artificial intelligence (AI) systems perform in the real world (Gebru et al. 2021; Marcus 2018). As Amershi et al. (2019b, p.298) put it, data science is 'all about the data that powers… [the] models.' Most research on data science, however, has focused on models or algorithms rather than the data itself (Hohman et al. 2020; Sambasivan et al. 2021). Data are viewed as being in service of the algorithms, an input 'processed' into abstract form and 'wrangled' to be explorable in statistical terms. In this view, data science activities are made sustainable by the efficient flow of data, which improves the predictions of algorithms. Better predictions create more value for users, who then become more incentivized to contribute further to the flow of data. At the center of sustaining this 'virtuous cycle' (Iansiti and Lakhani 2020, Ch. 3) between data and algorithms has been data scientists, as the ones who possess the ability to use tools for wrangling data and building models and algorithms.

Recent CSCW and HCI research implies that this virtuous cycle between data and algorithms does not fully capture challenges to sustaining data science activities 'in the wild' — in the practices and settings of domain experts where models are to be actually used (e.g., Stadelmann et al. 2018; Beede et al. 2020; Passi and Sengers 2020). The difficulty of collecting and managing data that are ultimately situated in domain experts' practices and settings commonly leads to data that are unreliable (inaccurately or inconsistently recorded) and incomplete (not enough columns and rows to capture the relevant phenomena). Unreliable, incomplete data create challenges for the sustainability not just in terms of the predictions made by algorithms (e.g., accuracy), but also in that predictions become less relevant to the phenomena of domain experts (Sambasivan et al. 2021; Beede et al. 2021; Gebru et al. 2021).

In this paper, we propose that sustainability in data science activities may be addressed by shifting the center of attention from models or algorithms to data and their situatedness in practice, and therefore from data scientists to domain experts. As a foundational theme in CSCW, sustainability has been examined in many ways, such as how humans adapt and appropriate technological artifacts beyond their original design and functions (e.g., Orlikowski 1992; Pipek 1995; Dourish 2003; Balka and Wagner 2006; Tscheligi et al. 2014; Muller et al. 2016; Meurer et al. 2018); the ongoing, contingent work of repair and maintenance (e.g., Orr 1996; Lutters and Ackerman 2002; Graham and Thrift 2007; Büscher et al. 2009; Jackson et al. 2012; Jackson 2014; Rosner and Ames 2014; Denis and Pontille 2015; Cohn 2016; Houston et al. 2016); and the long-term collaborative technology use and design (e.g., Dourish et al. 1996; Karasti et al. 2010; Borgman et al. 2012; Pipek et al. 2017; Ludwig et al. 2018).

Our proposal is inspired by CSCW and HCI research on the end-user design of software tools, which has long viewed sustainability as driven by making domain experts the ultimate 'owners of problems' (Fischer et al. 2004, p.35). This view continues in recent research to support domain experts to explore and solve problems using technical data science tools (Amershi et al. 2014; Gil et al. 2019). We seek to extend these domain expert-driven perspectives to the issue of sustainability in data science. Our proposal is that domain experts, as the ones who often collect and use the data in their practices and settings, should be viewed as not just owners of problems, but 'owners of data.' An 'owners of data' view emphasizes that sustaining the flow of data needed for data science activities ultimately depends on domain experts' situated data practices.

One example that points to domain experts' central role in sustainable data science activities is the idea from database design of 'user-defined data types' (Stonebraker 1989). When relational database systems were first being widely diffused to organizations in the 1980s, database designers found it unsustainable to expect domain experts to inform technical workers of how to define the data in these systems. Definitions would end up continually changing and expanding as the domain experts engaged with their data. The solution was to design database tools to enable domain experts to flexibly define their own data according to their own practices. For example, investment managers might define 'months' not based on a calendar, but as uniform 30-day increments to compare returns over periods of equal duration (*ibid*.). Despite growing attention to domain experts in CSCW and HCI research on data science, however, the database field's long focus on domain experts' data practices has remained largely invisible in this research.

Insight into domain experts' data practices would shed light on challenges for the sustainability of data science activities 'in the wild' by surfacing how their data practices mediate the interplay of data and algorithms. Such insight could inform the design of data science tools that support domain experts to collect reliable, complete data in their practices and settings, and to help make predictions or other results of models that use these data relevant for them.

To develop insight into domain experts' data practices and how they mediate the sustainability of data science activities, we conducted ethnographic fieldwork at a craft brewery in Korea. The brewery provided an ideal site for studying domain experts' data practices and their implications for sustainability. Craft brewing is an example of craftwork, which relies on the ability to engage with unpredictable materials and settings (Becker 1978; Sennett 2008; Marchand 2016). Craftwork is therefore less amenable to a more industrial view of data science based on statistical predictions using abstract data. Craft brewers' collection and use of a variety of data is critical to controlling the quality and consistency of their beers (Fisher et al. 2020; Bamforth 2003). The data (e.g., temperature, pH, and carbonation readings) are situated in that it needs to be continually measured, interpreted, and acted on at multiple stages of the brewing process, and in the brewery itself.

In our study, we first studied challenges in how the brewers collected and used data at the brewery, identifying several situations where their data practices led to unreliable, incomplete data: (i) variables defined but data not collected; (ii) data on variables collected but not reliable; and (iii) unclear and changing variables. To study how these challenges mediated the sustainability of data science activities, we then led a pilot project to build machine learning (ML) models for predicting yield efficiency (the amount of beer produced relative to the grain used) at the brewery. Analysis of the project showed how unreliable, incomplete brewing data created challenges for (i) wrangling the data *post hoc*; (ii) making sense of the ML model predictions from a technical perspective; and (iii) making the ML model predictions actionable in the brewery.

Our findings contribute to data science research in CSCW and HCI by identifying challenges for the sustainability of data science activities in terms of domain experts' data practices. First, the brewers used primarily logical relations in their (situated) data that were challenging to align with the statistical use of (abstract) data by data scientists. Second, the brewers found the value of reliable, complete data unclear since they could not see the value of data science tools. Third, the reliability and completeness of the brewing data depended on how the brewers collected and managed the data at the brewery — there were severe limits to what could be done with *post-hoc* activities (e.g., wrangling). To address these challenges for sustainability, our study also contributes design implications. Going beyond support for making predictions, data science tools could support domain experts' different uses of data, enable domain experts to see the value of data science tools, and to flexibly manage data as they are collected.

Our main theoretical contribution is a proposal for sustainability for data science activities driven by domain experts. In the 'owners of data' view that we propose, the central role shifts from the data scientists to the domain experts. The cycle that sustains data science activities is extended from the interplay between data and algorithms, to include how domain experts' data practices mediate the interplay between the two.

The paper is structured as follows. The next section offers theoretical background, framing our emphasis on domain experts' data practices within CSCW and HCI research on the situatedness of data and work practices. We then describe our methods. After that, we illustrate three situations where brewers' everyday data practices led to reliable, complete data, and how the brewing data created challenges for building ML models. We then discuss how our findings led to our proposal for viewing domain experts as 'owners of data,' and lay out challenges of aligning domain experts' data practices with data science activities. We conclude with limitations and future directions for research.

4.2 Background

4.2.1 The data in data science

Data science research has given surprisingly little attention to data, focusing more on the development of models or algorithms 'independent of which dataset is used' (Hohman et al. 2020, p.1; Sambasivan et al. 2021; Wagstaf 2012). The work underlying data practices has been often referred to as the 'grunt work' or 'janitor work' of data science (Lohr 2014). Practices of creating, collecting, and managing data tend to be seen as tedious and uninsightful tasks necessary for transforming unstructured data into datasets suitable for statistical models (Fletcher et al. 2020). Although there is growing interest in HCI, STS, and CSCW in data as a site to understand our relationships with technology (e.g., Ribes and Jackson 2013; Pine and Liboiron 2015; Feinberg

2017; Sanches and Brown 2018; Scheuerman et al. 2021), data science research itself still projects a view of data practices largely as a means to increase model performance (e.g., Sun et al. 2017).

4.2.2 'Wrangling' unreliable, incomplete data

There is broad agreement that data in real-world data science activities are often unreliable (not accurately or consistently recorded (e.g., 'messy') and incomplete (e.g., missing variables or values such that there are not enough columns or rows of data for the model) (Muller et al. 2019a; Zhang et al. 2020). Data typically fail to meet basic standards of reliability and completeness for several reasons. Organizations may not know what data to collect or how (Redman 2018). Documentation and guidelines for managing data and facilitating collaboration are typically lacking (Gebru et al. 2021; Mitchell et al. 2019; Sambasivan et al. 2021; Boyd 2021). Unreliable, incomplete data lead to time-consuming 'data wrangling' tasks (e.g., curating, cleaning, formatting) and, more seriously, to 'data cascades,' or 'compounding events causing negative, downstream effects' on model implementation (Sambasivan et al. 2021, p.5). Research in HCI and CSCW has started to examine such implementation problems across settings such as healthcare (Sambasivan et al. 2021; Beede et al. 2020), criminal justice (Chancellor et al. 2019), and finance (Passi and Barocas 2019).

A variety of technical and organizational proposals have been advanced to support tasks of wrangling missing or messy data, and to ensure (in a statistical sense) the quality of the wrangled data. Proposals include developing interactive visualization methods to aid data transformation (Kandel et al. 2011), automating data inefficiency detection and repair (Hynes et al. 2017; Krishnan et al. 2016), incorporating the context of data's domain (e.g., master data, reference data) into wrangling tools (Koehler et al. 2017), and better aligning data wrangling tools and data scientists' data exploration workflows (Drosos et al. 2020).

These extant proposals for data wrangling, however, remain mostly *post hoc* and abstract. The support is for improving the reliability and completeness of data only after the data have been collected (e.g., Pine and Liboiron 2015). There is little regard for how the data and the practices for collecting them were situated in practices and settings. A motivation underlying our study was to ask the question of what makes data wrangling tasks necessary in the first place: Could domain experts' situated practices of creating, collecting, and otherwise engaging with data be leveraged to improve the reliability and completeness of data, and with less reliance on tedious *post-hoc* wrangling?

4.2.3 Situating the data in data science

A view of data and data practices as situated is well-established in CSCW, HCI, and STS. Critical data studies, for example, stress that the meaning and function of data are not given and abstract but determined in the context of its production and use. These studies stress the central role of processes and negotiations involved in making data work at the site of practice (e.g., Bowker 2005; Borgman 2016; Gitelman 2013; Taylor et al. 2015; Feinberg 2017; Bossen et al. 2019; Pine and Liboiron 2015; Scheuerman et al. 2021). The emerging field of human-centered data science (HCDS) has surfaced a variety of human activities and perceptions critical to data science projects (Muller et al. 2019b, Kogan et al. 2020; Aragon et al. 2016; Passi and Jackson 2017). Muller et al. (2019a) draw attention to how data science workers engage with data, analyzing the workflows of data science workers in terms of data 'interventions' such as discovery, capture, curation, design, and creation. Wang et al. (2019b; and also Wang et al. 2021) explore how data scientists perceive automated AI systems and how the introduction of such systems may change how data science workers work with data. Several studies examine possibilities of integrating human-driven qualitative methods with ML methods (e.g., Baumer et al. 2017).

A number of studies in HCI and CSCW have also begun to explore how data science projects 'in the wild' are 'made to work' (Stadelmann et al. 2018; Passi and Sengers 2020). Based on ethnographic fieldwork in a data science team within a US corporation, Passi and Jackson (2018) show how data science projects involved processes of dealing with different goals, expectations, and skills between managers, business analysts, and data scientists to establish shared confidence and trust in the workings of data science models. Beede et al. (2020) provide an account of an AI-assisted eye-screen system deployed in clinics in Thailand. They show how the performance and effectiveness of the system was not determined by its predictive accuracy alone, but depended on socio-environmental factors at the physical sites (e.g., poor lighting led to ungradable images; internet connectivity issues caused delays for patients). These studies emphasize that designing and deploying data science tools is a multi-stakeholder endeavor that requires ongoing, situated work to 'align system working with shifting expectations' (Passi and Sengers 2020, p.2) and with the material conditions of the site of deployment.

While this prior research attends to the need to negotiate between and align different values, expectations, and resources surrounding data science activities, an underexplored question is how to sustainably align the work of data scientists with the data practices of domain experts who initially create, collect, and engage with data in their settings. How do domain experts' data practices shape the effectiveness of data science activities, and what does their relationship imply for the design of sustainable data science tools?

4.2.4 Bringing domain experts into data science activities

The role of data scientist itself is still a subject of debate (Muller et al. 2019a; Chatfield et al. 2014). Several scholars have proposed a broader term of 'data science worker' to refer to anyone for whom data-related activities are critical to their work to examine a wider range of roles involved in data science (e.g., Muller et al. 2019a; Zhang et al. 2020). Other studies have examined domain experts as 'data workers' who closely engage with data and data science activities to inform their decisions, but who would not identify themselves as data scientists (Boukhelifa et al. 2017; Liu et al. 2020). While domain experts' level of technical skills and engagement with data science applications may vary, data scientists and domain experts operate in different worlds (a world of models vs. real-world settings of practice) (Viaene 2013), with the former most often taking up the role of building and engaging with technical data science applications.

Of particular relevance to our interests in sustainability are studies of data science 'in the wild.' These studies point to possible tensions between the goal of data science as a 'universal science' (i.e., in the work of technical data scientists) and applicability to particular domains (i.e., what domain experts know and how they work) (Ribes 2019, p.524). While data science is often described as 'the study of the generalizable extraction of knowledge from data' (Dhar 2013, p.64), given that data themselves are situated in a particular domain of practice, data science tools need to be made to work for domain experts as well (Rudin 2019).

Prior research in this area has focused on reducing potential tensions between statistical generalizability and domain-specificity. Examples include facilitating collaboration between domain experts and data scientists by building common ground or shared mental models (Mao et al. 2019; Piorkowski et al. 2021), structuring the stages of collaboration in a data science workflow (Zhang et al. 2020), or capturing or incorporating domain experts' knowledge for tasks such as data labeling building models, or contextualizing model results (Muller et al. 2021; Aslan et al. 2017; Kery et al. 2018; Zhang et al. 2020; Carter and Scholler 2016).

4.2.5 Sustaining data science activities by domain experts as the 'owners of data'

Despite the growing interest in collaboration shown in the above research, domain experts' role in data science tends to be limited to informing the data scientists (e.g., providing data, informing the data scientists of their domain-specific knowledge, or giving feedback on model results) (Amershi et al. 2014; Kery et al. 2018; Gil et al. 2019; Muller et al. 2019a; Hohman et al. 2020; Zhang et al. 2020). A restricted role of domain experts (e.g., to informers of data scientists) may lead to inefficient and ineffective collaboration (Amershi et al. 2014; Gil et al. 2019) that impedes the sustainability of data science activities in several ways.

First, while domain experts are the ones who usually create and collect data, existing research tends to assume data as already collected and given. This assumption does not account for how mismatches arise in how the data are used in statistical models versus how they are used by the domain experts in their practices and settings (Gebru et al. 2021; Passi and Barocas 2019; Gil et al. 2019). As Sambasivan et al. (2021) argue, the emphasis of statistical modeling is to fit the model to the given data, rather than to evaluate whether the data are relevant to the problems in domain experts' work practices.

Second, the technical work of data science projects is typically approached as the 'one-off application' of a statistical model to a given dataset (Polyzotis et al. 2017). Built on an assumption of a 'largely stable world' (Marcus 2018), data science often views changes to data (or its underlying distribution), called 'data drift', as detrimental to model performance (Hohman et al. 2020; Hoens et al. 2012; Williamson and Henderson 2021). For data science activities to be sustainable, data need to be managed to account for changes to data in the dynamic world (e.g., Amershi et al. 2019b; Bopp et al. 2017). While several studies explore ways to set up a 'pipeline' for data science activities with continuously incoming, and possibly changing, datasets (e.g., Roh et al. 2021; Breck et al. 2019; Lourenço et al. 2019), these activities mostly center on data scientists or data science workers. These studies have little to say about how such datasets are initially produced or shaped by domain experts. Domain experts should be well-suited to contribute to setting up and managing data pipelines since they commonly are the ones who create and collect data according to their (shifting) work settings and practices.

A growing number of studies have begun exploring ways to deepen domain experts' involvement with data science tools as a type of end-user (Amershi et al. 2019a). Insights have been developed, for example, into how domain experts may themselves explore and experiment with their data and model settings (Seidelin et al. 2020; Ferreira and Monteiro 2020), and into how providing more interactive visualization techniques and interface design may allow domain experts to directly build and modify models relevant to the context of their data and problems (Amershi et al. 2014; Gil et al. 2019).

Building on these studies, and on our ethnographic fieldwork at a craft brewery in Korea, we explore ways to deepen domain experts' involvement in data science as the 'owners of data'. By studying domain experts' data practices and how these practices mediate data science activities, we seek to develop insight into challenges of sustainability in developing and using data science tools, including possibilities for design support.

4.3 Methods and analysis

4.3.1 TheBrewery

This paper builds on ongoing fieldwork at a craft beer brewery in Korea that was founded in 2016, hence referred to as TheBrewery. The original head brewer (HB1; see Table 2) left the year prior to our study. During our fieldwork, it was operated by four male brewers in their 30s and 40s (B1-B4). Operations were overseen by a team of four managers in Seoul, several hours' drive from the brewery, which was located in a tourist area). The CEO (M1, male, 40s) and founder

(M2, male, 50s) are featured in our study. During the first part of our fieldwork, a male intern in his 20s worked with the brewers; he left after his one-year internship. The brewers communicated frequently with the management team via email, phone, and messaging apps, as well as in person. We summarize the brewery members who appeared in this study in [**Table 2**] below. While the production facility was of average size for a craft brewery in Korea, the beer had developed a strong brand and reputation for quality and innovation. The beer was distributed in bottles or kegs across the country (e.g., in convenience stores, supermarkets, and trendy restaurants), and the brewery was profitable and growing.

Table 2. Members of TheBrewery featured in chapter 4

Role (code)	**Age**	**Nationality**	**Experience at/with the brewery**	**Notes**
Brewer (B1)	30s	Korean	3.5 years	Takes an informal leadership role.
Brewer (B2)	30s	Korean	3 years	
Brewer (B3)	30s	Korean	3 years	
Brewer (B4)	40s	American	2 years	
Former head brewer (HB1)	30s	Australian	3 years	Left the brewery the previous year.
CEO (M1)	40s	Korean	2 years	
Founder (M2)	50s	Korean	4 years	Initiated project for 'systemization' of brewery operations focused on digitizing data and integrating AI (see 4.3.2).

Craft brewing offered a favorable site for studying domain experts' data practices and how they mediated data science activities. Craft brewing is an example of craftwork, which relies on the ability to engage with unpredictable materials and settings (Becker 1978; Bamforth 2003).

Craftwork is therefore less amenable to a more industrial view of data science based on statistical predictions using abstract data. Craft brewers are also known to be ‘data geeks’ enthusiastic to measure all sorts of variables (e.g., constantly monitoring temperature, humidity, pH, ABV (alcohol by volume), etc.) to improve quality and experiment with new recipes or techniques (Morgan et al. 2020). At the same time, craft brewers have only begun to show interest in data science and ML (e.g., Gonzalez Viejo et al. 2019; Fisher et al. 2020). The craft brewers’ rich existing data practices enabled us to study challenges in making novel data science tools actionable within these practices.

Our site was also favorable to our research question in that brewing in general is a complex, multistage process where control naturally involves collecting a large variety of data variables. Brewing depends on manipulating grains (typically barley malt) at precise temperatures, and controlling a live fermentation over weeks (Hough 1991; Lima et al. 2011). The ingredients are continually in flux across many points in time. Data collection in a craft brewery is especially challenging, as a defining characteristic is dealing with a much higher variety of specialty ingredients (e.g., malts, hops, yeasts) and recipes (e.g., seasonals, specialty ales, etc.) than an industrial brewery (Morgan et al. 2020). Studying challenges in craft brewers’ data practices could point to ways in which data work can be aligned with the work of data science.

4.3.2 The systemization project

TheBrewery had been realizing strong growth and profitability. Korea’s craft beer market was also growing (Seo 2018). During the beginning phase of our fieldwork, M2 informed the management team and brewers that the timing was right to embark on what he called ‘systemization’ to prepare for a large expansion in its production capacity. In a meeting with M2 and the four brewers during our first week of observation, M2 articulated that expansion would mean both introducing new equipment to the existing brewery facility, as well as opening additional facilities. He explained that, by 'systemization', he meant systematically collecting and analyzing data to develop quality control standards and to monitor the production process across their variety of recipes amid expansion.

4.3.2.1 Data needed for systemization

Systematization was perceived as necessary by the brewers, as well to enable TheBrewery to maintain consistency amid changes in brewing practices that would take place from expansion regarding:

- Facility size (B1: 'If [we can use data to make] our work… standardized and consistent… we will get similar outcomes and beers even as the brewery gets bigger.')
- Turnover of brewers (B4: '[we need] to accumulate data for consistency in the quality of our beer, regardless of who the brewers are.')
- Equipment (B2: 'If the brewery gets bigger, we may have bigger equipment. I need to know whether the result that I got when I used equipment sized A will be similar to the result [next time] with equipment sized A100? Otherwise, if something goes wrong, we won't even know what specific [corrective] action we could take. [If we have data], we [the brewing team] could together review and evaluate our practices, and talk about how we could make our practices more consistent.')

4.3.2.2 Issues with current data practices

The management and brewing teams all acknowledged that a core challenge to analyzing data for systematization would be overhauling the current practices and software for collecting and managing data at TheBrewery. Data on the brewing process currently were collected on 149 variables across three paper brew sheets, comprising:

- Brew logs (85 variables (recorded on a separate brew sheet for each batch (½ day process), with each 'brew' (the basic unit of production) consisted of one to three batches) on the process of turning barley malt into a fermentable liquid, or 'wort'),
- Fermentation logs (16 variables (typically recorded daily, for two to three weeks), on the process of fermenting the wort into beer), and
- Packaging logs (48 variables on putting fermented beer into bottles or kegs (as part of their branding strategy, they did not use cans)).

Variables included: the duration of steps, temperature and pH readings, readings of the gravity (a measure of the beer's density of fermentable sugars), pH, volume, dissolved oxygen (DO), and carbon dioxide (CO2) levels.

Brew sheets were designed by the former head brewer (HB1) as Excel templates, and customized for each recipe. Each template came with some boxes filled in with preset values, corresponding to the recipe standard or theoretical values based on brewing science. The brewers would commonly discuss these preset values prior to each batch to make adjustments versus the previous batch, though most of the preset values would typically stay untouched. The brewers would print out a set of brew sheets (one to three brew logs depending on how many batches would go into a single brew, one fermentation log, and one packaging log). The printed brew sheets would be inserted into clipboards, which would follow a brew around the facility as it progressed (e.g., from the boiling tanks, to the fermentation tanks, to the packaging area).

The most basic change to the data practices that the members of TheBrewery sought was simply to digitize the data from the paper brew sheets. Currently, the paper brew sheets were filed away in plastic binders once a brew was no longer active (i.e., when the finished beer was packaged into bottles or kegs). In order to analyze patterns in their prior brew sheet data, the brewers relied primarily on physically retrieving the paper brew sheets. The brewers had initially asked the intern to transfer the brew log data (measurements, not comments or annotations) to an Excel file. The data were only partially transferred, however, as the amount of time required for data entry was much larger than they had anticipated, and the file was hardly used. The members of TheBrewery expressed the need for software to facilitate or even automate the collection of brewing data.

4.3.3 The yield efficiency problem

About a year into our fieldwork at TheBrewery, one morning M2 expressed to us his eagerness to begin a pilot project to develop ML models for analyzing the brewing data 'systematically'. We said to M2 that we would be able to develop a pilot ML model, provided there would be no monetary compensation and that we could collect field data on the process of building the model for the purposes of doing research. We believed that, given the field data that we had already collected on the brewers' data practices, studying the process of building an ML model at TheBrewery would enable rich observation of the relationship between the brewers as domain experts, their data practices, and the use of data science tools.

M2 was enthusiastic and arranged a meeting and lunch with the four brewers the following week, and a subsequent meeting with the management team in Seoul to develop a specific problem and timeline for the pilot project. In conversations with the brewers and an outside consultant (the third author of this paper) with 14 years of operations experience in IT project management as well as ML expertise, it was agreed to focus on analyzing the brewing data to gain insights into 'yield efficiency', or the efficiency with which the barley malt (the main ingredient) is converted into packaged beer. Yield efficiency is notoriously volatile in craft beer breweries (Bamforth 2003). It is a valuable problem to address in that low or unpredictable efficiency can cost breweries large amounts of money in extra ingredient costs and the need to buffer inventories due to uncertainty about how much beer will be produced from a batch. Next, we describe how we collected field data both before starting the pilot project (on the data practices of the brewers), and for the process of building the ML models.

4.3.4 Data collection

4.3.4.1 Fieldwork: observation, brew sheets analysis, and workshops

Our findings on the data practices of the brewers (section 4.4.2) derive from field data collected by the first and second authors, which took place over a period of 1.5 years. They were introduced to the brewers by the brewery founder (M2) and explained the interest in studying the brewers' data practices (e.g., how they worked with data during brewing, and how they collected and maintained the data). We noted that our field data would be first reviewed with the brewers and also that it may be used in published research. After the brewers agreed to participate in our research, the first and second authors began to collect field data through participant observation, data artifacts, workshops, and interviews.

First, prior to going into the site, the first and second authors held exploratory semi-structured interviews with the brewery's management team (M1 and M2) to familiarize themselves with the brewery's data practices and goals relevant to its use of data. In addition to M1 and M2, we interviewed the marketing manager and distribution manager in person as well as the former head brewer (HB1) via WhatsApp, with a focus on the value of data at the brewery and how data were collected and maintained.

The first and second authors made seven trips to the brewery over 1.5 years of fieldwork, with stays between several days to two weeks. In our first visits, the emphasis was on familiarizing ourselves with the brewing team and their practices. With permission, we shadowed the brewers during their work practices and took part in other activities (e.g., sharing meals and snacks, helping organize their storage room, participating in their discussions and meetings). During our visits, we had numerous informal exchanges with the brewers and the intern, took down notes, and photographed and video-recorded both their brewing actions and how they collected and used data in the brew sheets, also with their permission.

To study their data practices over time, we digitally scanned 1,200+ paper brew sheets that had been stored in plastic binders. Going over the different formats, annotations, and data values on these sheets enabled us to see patterns in both their brewing and data practices. The brew sheets also served as a helpful memory aid to get the brewers to talk about their brewing and data practices based on their experience.

After scanning all the brew sheets into digital files, we manually put together a list of all data variables (149+ variables), based on which we conducted four 'data workshops' with the brewing team at their office and on Zoom (fourth workshop). In these workshops, we used a projection screen to go through one by one all the data variables from the brew sheets grouped by production stage. This exercise prompted lively discussions with the brewers on (i) how the variables related to their work processes (first workshop); (ii) the meaning and importance of the variables in the context of the pilot project on building an ML model of yield efficiency (second workshop); and (iii) potential challenges for the introduction of data science tools (third workshop), especially regarding the implications of the results of our ML models for predicting yield efficiency for data science practices at the brewery (fourth workshop). The workshops lasted between 60 minutes and two hours.

4.3.4.2 Pilot project for building ML models

Our findings on the process of building ML models of yield efficiency at TheBrewery (section 4.4.1) derive from the second and third authors' participation in the ML pilot project. The second author had a background in database design, and handled the building of the dataset for the

ML models. First, he put together a dataset on yield efficiency-related variables from the brew sheets. This process took place (non-continuously) over a period of 10 months, beginning with digitizing the brew sheets with a scanner, and having the data in the brew sheets manually entered into a Google Sheet with several tabs. He then "wrangled" the data into a single spreadsheet to prepare it for the ML models (columns irrelevant to the ML pilot project (see 4.3.3.2.2) were removed; the rows of data for the remaining columns were cleaned to correct or impute values). The third author had specialized knowledge in ML, and handled the coding and technical evaluation of the ML models for predicting yield efficiency.

The dataset. A first step to building the ML model was to identify a measure for yield. The second author labeled the digitally scanned data and had an assistant manually enter the data values in a Google Sheet, with separate tabs for the brew log, fermentation log, and packaging log (a scanning app for automating data entry was tried, but was not reliable). After all the data values were entered and columns with no values were removed, there were 289 observations in the brew log and 85 variables, 172 observations and 10 variables (over 21 timestamps (days)) in the fermentation log (there were fewer observations since one fermentation could contain multiple batches); and 124 observations and 46 variables in the packaging log (there were fewer packaging observations since one packaging event could involve multiple fermentations).

A major concern that arose once the data were entered was the large number of missing values, in particular yield-related values (e.g., 'volume transferred', a measure of how much liquid was transferred from one stage of brewing to the next), as well as suspect values in many places (e.g., wildly different values for pH in the same batch). 29.9% of the brew sheet values were not recorded, 18.8% for the brew log, 38.2% for the fermentation log, and 53.6% for the packaging log. For 27% of the variables included on the brew sheets, more than half of the data were missing.

Despite concerns about the reliability and completeness of the dataset, the authors decided to go ahead with the pilot project. For one, M2 had originally wanted the project to be an exploratory learning experience for the brewers. There would be potential benefit to going through the process, and there was no hard expectation that the predictions from the ML models would be immediately implemented. The process could surface specific ways in which the brewery could improve its data practices. The authors were themselves curious how the unreliable, incomplete data would exactly affect the actionability of the ML models. Going through the pilot project could give both theoretical and practical insights for how the brewers' data practices and use of ML models or other data science tools could be aligned.

Building ML models. The third author wrote the code for the models, based on the dataset provided by the second author. As the second author had also done fieldwork with the first author at TheBrewery (and had done several years of prior fieldwork for another project at several other brewing facilities), we agreed that he would play the role of a proxy 'domain expert', informing the third author (who had never visited the site or talked to the brewers) about the relevance of the brewing data variables (features) for the ML models. The rationale was to roughly simulate the process of the actual domain experts (the brewers) interacting with a data scientist. The second and third authors thus jointly selected subsets of 'features' for the ML models from the brewing dataset that they suspected would be relevant to yield efficiency at the brewery.

The iterative process of writing code was divided into eight 'sprints', or one- to two-week iterations of versions of the ML models. At an initial planning meeting (conducted on Zoom), the second and third authors agreed that the sprints could focus on predicting yield based on the variable 'mash percent' (a measure of yield at the stage of brewing where malt is initially mixed with water). The second author noted from his fieldwork at the brewery that the mash process was critical to yield. The authors noticed further that, compared to other plausible outcome variables in the dataset, 'mash percent' had fewer missing values. They also agreed that the initial sprint could draw on 47 features from the brewing data that had plausible relationships to yield (based on the workshops conducted with the brewers (see 4.3.3.1)).

For each of the sprints, the third author ran a baseline linear regression model (LM), and four different ML models (generalized Bayesian model (GBM), random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN)). The simpler linear model was to test whether it was worth using any ML model, as all of the ML models assumed more complex non-linear relationships among variables. Four ML models were used since it was not initially clear which type of model would make the best predictions (trying out multiple models is the common approach in using ML in practice).

After the initial planning meeting, the two authors conducted eight sprints over a period of 52 days. The structure of each sprint was as follows. First, the second author would give a set of features to the third author. The third author would return a 'sprint report' in a PDF, composed of (i) a comparative graphic of the quality of the ML predictions (including the non-ML linear model), depicting the R and R squared values (measures of the percent of variance explained by each model); (ii) a linear correlation matrix and heat map on all the current features in the model; (iii) five visualizations of the nonlinear behavior individual features based on explainable AI (XAI) — a recently emerging set of techniques for evaluating the relevance of features specific to ML models (Arrieta et al. 2020). The third author also added comments on how the results looked from a technical data science perspective (e.g., an abnormal bimodal distribution in the data). Second, the second author returned a list of features to remove or change, based on using his familiarity with the brewery from his earlier fieldwork. Third, the two authors communicated (over a messaging app, in person, or over a Zoom call) to confirm the changes to the features.

Since the idea of the sprints was not just to build a good predictive model, but to study the domain-specific process of building ML models, the two authors created a shared Evernote document for each sprint, in which they wrote general notes and specific comments for each feature that was removed or modified. They stored the sprint reports, comments on each report, and other files and reference materials. Throughout the sprints, through Evernote and over messaging apps, Zoom meetings (30 minutes-1hour), and an in-person 'workshop' at a cafe (2.5 hours), the third author provided materials and explanations to the second author on ML and XAI techniques; the second author shared insights from his field study of the brewery and brewing science in general.

4.3.5 Analysis

Analysis of data from both the fieldwork and the ML pilot project — spanning interviews, participant observation notes, videos, audio files, photographs, the brew sheets and other brewing-related documents, online chat logs, and the Google Sheet dataset, ML code, and sprint reports — was conducted by the first and second authors. All interviews, videos, and audio recordings in Korean were transcribed and translated into English by the first author, who is a native Korean speaker. Other materials in English were transcribed by the first and second authors. Transcripts were anonymized and edited lightly for readability.

Our data collection and analysis for this current study happened in three rounds. First, since the initial goal of our study was to provide guidance on the brewery's planned 'systemization' project (see 4.3.2), we needed to gain insight into how and why the brewers were collecting and using data at the brewery. Adopting a situational analysis approach (Clarke et al. 2015), transcripts and notes from the fieldwork were open-coded to surface possible themes, which were then discussed and iteratively modified. Subsequent field visits and interviews led to further discussions and modifications as we studied the brewers' data practices in progressively greater detail. Through situational analysis, our initial fieldwork mapped out relationships among the brewers, management team, and non-human actors (e.g., the brew sheets, ingredients, and equipment in the physical setting of the brewery), as well as how broader social, cultural, and economic factors (e.g., nature of conversations among the brewers, their habits of using paper brew sheets and not digitized data, different goals and values associated with data collection) could affect the brewer's data practices. The findings from this round of analysis were shared and discussed with the brewers in our third workshop.

The second round of data collection and analysis focused on the ML pilot project. Here, too, we used a situational analysis approach to understand the relationship among human and non-human actors (e.g., the data scientist, proxy domain expert, digitized brew sheet data, ML code, and visualization techniques) and their relationships and positionalities (e.g., the notion of 'good enough' data or models in a statistical sense versus from a practical/operational sense). The ML model results were then shared and discussed with the brewers in our fourth workshop.

The final round of analysis concerned integrating the themes that emerged from the two settings (the brewery and its data practices, and the building of the ML models), given the brewery's goal of exploring possibilities for integrating data science tools into the brewers' work practices. Also drawing on our additional data (data from the fourth workshop and confirmatory interviews with the brewers and management), we tried to map out relationships among the themes that we identified. We arrived at the view that the reliability and completeness of data collected by the brewers could be the central theme for illustrating the connection between the brewers' data practices (i.e., how they result in data quality issues) and the effects of these practices on the actionability of the results of the ML models as data science tools (i.e., how the actionability of the model predictions was affected by data quality issues). We believed that this theme would be also useful for deriving concrete design implications for data science tools that can be used by domain experts as part of their work practices.

4.4 Results

4.4.1 Data-related challenges in building the ML model were rooted in the brewers' data practices

In this section, we examine data practices at the brewery, surfacing three types of situations that led to unreliable and incomplete data: (i) variables defined, but data not collected; (ii) data on variables collected, but not accurately or consistently; and (iii) variables not defined clearly or subject to change. Based on both observations and brewers' reflections on these situations, we show how unreliable, incomplete data arose from material (e.g., issues with data entry tools, changes to brewing practices and settings), organizational (e.g., different values of data, lack of standard operating procedures), and knowledge-related (e.g., seeing theoretical values as data, reasoning with data in logical terms) factors.

4.4.1.1 Variables defined, but data not collected

As mentioned in 4.3.2.2, almost 30% of the values on the brew sheets were left blank, and more than half the data were missing for 27% of the variables. Data on these variables were not collected for a variety of reasons.

Lack of resources for data collection. The brewers did not measure certain variables included on the brew sheets because they did not have the equipment or manpower. The brewers explained that variables such as 'post-boil volume' (liters of liquid after boiling the malt) and ‘post-boil plato’ (density of sugars after boiling the malt) were not collected because the mash tun (a large stainless-steel container for boiling the malt) would get too hot to directly measure them safely with the thermometers that they had on hand. In addition, a grid at the bottom of the packaging log contained seven quality control variables (plato (density of sugars), pH, WLD (a test for yeast contamination), HLP (a test for bacterial contamination), CO2 (percent carbon dioxide), DO (percent dissolved oxygen), and TPO (total concentration of oxygen)), to be measured across seven stages of the brewing and packaging process, for a total of 38 values. None of these data was recorded in any of the batches. The former head brewer (HB1), who designed the brew sheet templates, commented, ‘Data that... pertain to quality... tends to be… measurable [only] if we had a nice lab’. A quality control program at the brewery still remained aspirational (e.g., B4: ‘we will start it one day’). The other brewers added that they not only lacked lab equipment for measuring these variables, but also lacked the experience and standard operating procedures for running the tests as well as the manpower to do so.

Some variables were not measured temporarily because of equipment failure. In the packaging logs, 40.4% percent of the measurements of two core quality control variables ‘CO2’ (the percent carbon dioxide in the beer) and ‘DO’ (the percent dissolved oxygen in the beer) were missing, mainly due to frequent breakdowns of the CO2 and DO meters. The brewery’s flow meter (a device for measuring the volume of wort during brewing) remained broken for the entire period (around 1.5 years) of our fieldwork. As a result, the variable ‘volume transferred’ — the volume in liters of the beer transferred from the mash tun (where the brewing stage of extracting sugars from malt happened) to a fermentation tank (where the next stage of adding yeast to ferment the sugars into beer happened) — which is the crucial outcome variable for measuring yield efficiency, was only indirectly estimated on the brew sheets, and only for 81% percent of batches.

Data collection not prioritized (lack of incentives for learning). The frequency of data collection was linked to the brewers' prioritization of data for learning about or controlling the brewing process. When variables included on the brew sheets were not measured due to missing or broken equipment, the purchase, repair, or replacement of equipment was deferred if the brewers did not feel that missing data would interrupt their work process or affect the quality of the beer.

The problem of yield efficiency in our ML models (see section 4.2), for example, was not directly related to the brewers' work processes or to beer quality, and so 'volume transferred' was not a high-priority variable. The brewers described the repair of the broken flow meter to measure the variable 'volume transferred' from the mash tun to a fermentation tank as 'extra work.' B4 commented: 'Flow meters aren't that expensive, but they break down easily and (measuring the volume transferred) is not one of our priorities at this point'. B2 added that they felt that they now had enough 'experiential data' on that variable, with which they could reasonably control the brewing process without actually measuring it: 'We just transfer (from the mash tun to a fermentation tank) everything we've brewed...We know that the levels would vary somewhat, but we don't think it's critical.' Instead of repairing or replacing the flow meter, some of the brewers recorded the variable based on just 'topping up' (adding water to a certain level of the fermentation tank) to estimate the volume based on how much water they had to add.

More broadly, the brewers commented that the data collection for brew logs was 'more important' than that for packaging logs since they saw the variables on brew logs as more directly related to their work as craft brewers. As mentioned in 4.3.3.2, packaging logs were most likely to have missing values of data (53.6%, versus 18.8% missing values for the brew log and 38.2% for the fermentation log) out of the three types of brew sheets. Many packaging logs only had data on variables such as '[packaged] date' (only one missing value out of 120), while leaving empty cells on other variables (e.g., ‘low’ (number of bottles or kegs filled low), ‘total volume,’ ‘tank volume,’ ‘loss volume’), which could be critical indicators for yield efficiency. The brewers conceded that they often recorded only the data they considered essential on packaging logs — that is, how my bottles or kegs were filled and when: ‘In the beginning, we did try to fill out the [packaging] log completely, but over time, I guess we just neglected the variables that we didn’t think were important to our work’ (B1).

Given that the needed equipment was in place, the perceived value of a variable was a more important factor in explaining missing data than the effort required for data collection about that variable. Two variables, for example, that all of the brewers considered of the highest importance for controlling their brewing process were ‘pH’ and ‘plato.’ ‘Normal’ readings of these variables (i.e., within their target ranges) were strong indicators that the brewing process was on track. While the brewers had to manually measure these variables (e.g., extract a sample in a beaker and perform multiple readings of its pH using a pH meter), the pH and plato variables had few missing values (e.g., only 10 missing values for the initial pH in the fermentation log out of 158 batches).

Data collection not integrated into brewing routines. Missing or incomplete data on existing variables also arose from a lack of integration of data collection activities into the brewers’ daily work practices. For example, six variables appeared on the bottom of the brew log for measuring properties of the yeast including yeast type, pitch time (time when yeast is added to start fermentation), and pitch rate (rate at which the yeast are added). Daily readings of these variables could be useful for understanding trends in yeast activity and potentially yield efficiency. While the brewery had the necessary equipment to measure these variables, no data had been recorded. HB1 explained the persistence of these missing values:

> ‘We could have routinized this stuff early on in the brewery, but introducing it now would be quite a change. So if I’m setting up a brewery, one lesson I take away is to start tracking core yeast activity variables right from the start.’

Integrating new data collection practices would thus be difficult to sustain.

Given broken or unavailable equipment for collecting data on variables that the brewers did not consider critical to their work, data collection also easily became un-routinized (such as the earlier example of the broken flow meter, in which the brewers got used to working with missing values for the variable ‘volume transferred’). The brewers phased out recording data on certain variables if they thought their experiential data on these variables seemed to provide good enough guidance on the brewing process. B3 commented on why he stopped recording the variable TPO (total oxygen) on the packaging log:

> ‘We’ve been collecting data (on it) for a while and it seemed that our beers tended to show values similar to or better than the target values. And for TPO values, we can kind of tell if they are off by looking at the amount of foam that forms during bottling.’

Here, data collection initially served to establish reference points for evaluating their beers relative to theoretical targets. Once the reference points had been established, collecting additional data on the variables became optional, leading to an increase in missing data. Data collection on variables whose values could be reasonably estimated with direct observations (i.e., ‘how much foam appears on top of beer’) was prone to discontinuation.

4.4.1.2 Data on variables collected, but not reliable

The previous section looked at situations at TheBrewery where, for various reasons, (lack of resources, lack of prioritization, and lack of integration into routines), data collection was incomplete. In this section, we look at situations where data were collected, but the values were known or suspected to be unreliable (e.g., inaccurate, inconsistent), identifying reasons for this unreliability.

Treating preset values as observational data. 34 of the 149 variables on the brew sheets had preset targets. The targets were determined from theoretical values from brewing science or from standardized recipes from brewing software. The preset targets were treated as observational data: the data were 'collected' in the sense that values were recorded on the brew sheets, but there were no actual measurements of these values.

The use of preset theoretical or standard values in place of actual data from the brewery stemmed from lack of measuring instruments at the brewery. For example, grain-related variables such as 'moisture' (percent moisture in the grain), 'diastatic power' (amount of enzymes available in the grains to convert starch to sugar), and 'protein' (percent protein) could not be directly measured at the brewery. The brewers used the data provided by grain suppliers to fill out the brew sheet. The brewery also did not have a spectrophotometer to measure the variable 'IBU' (International Bitterness Units). The brewers instead used recipe-based formulas to calculate theoretical values of IBU and recorded these values on their brew sheets.

The brewers also developed a habit of pre-writing data on temporal variables (53 of 149 variables) of the brewing process, such as 'boil time' (time boiling the malt) (usually one hour), 'whirlpool time' (time for stirring the boiled malt to remove hop and malt particles) and 'cooling time' (time for cooling the boiled malt to fermentation temperature) (usually five minutes). B4 recounted one such experience regarding data on 'boil time':

> 'Sometimes we all just pre-write because we know the time - alright, we should do it then. And then something happens and I'm like, "ah, I already wrote it, I'm not gonna erase it"…Yesterday we were brewing something new, and my boil time was longer than it usually is. I would've missed that because normally I just pre-write and…don't really look (at the boil time on the brew sheet).'

For the brewers, the data on these variables served as a check on which action to take next, rather than a check on if actual values diverged and corrective actions needed to be taken.

From the perspective of collecting data for the ML models, the use of preset data in place of actual measurements had led to severe reliability issues (see 4.4.2.1). The brewers commented, however, that while they had the intention to measure the variables accurately, they also did not believe that the preset values would diverge from the actual values to an extent that the difference would noticeably deteriorate the taste of beer: 'Even with all these differences…I think the taste of our beer has always been somewhat consistent…If you are really sensitive, you might notice that [a beer] is less sweet, or sweeter, than usual, but that's about it' (B3).

Cascading differences between actual and preset values. When actual data on variables differed from preset values already recorded on the brew sheets, the brewers typically did not make changes to the values on the brew sheets. Instead, the brewers made changes to other interdependent variables in the brewing process. B2, for example, explained how the brewers would make adjustments to divergent measures for the variables 'strike volume' (amount of water when adding the malt; usually set as 1300L) and 'sparge volume' (amount of water when draining the spent malt; usually at as 1200L):

> 'There could be more strike water (>1300L) depending on the condition of the machinery but we don't know the exact amount. If we seem to have used more strike water, then we try to use less sparge water (<1200L) in the following stage. What we are trying to keep constant is the sum of the amounts of strike water and sparge water.'

Similarly, B1 and B2 explained how they aimed to make on-the-spot ballpark adjustments to 'pre-boil volume' (amount of water after draining the malt, and prior to boiling the liquid) and related variables based on their preset values and interdependencies with other variables:

> B1: '"Pre-boil volume" is given (e.g., 2000L)…We add "sparge volume" to "mash volume" [amount of water after malt is added]...(as) we add around 1L of water per 1kg of malt…then "pre-boil volume" is roughly equal to "mash volume" + "sparge volume" - "grain weight"...according to which we could adjust the amount of water and ingredients.'
> B2: 'That kind of makes sense conceptually and we make adjustments loosely based on that, but I'm not sure if we could formulate (the relationship) mathematically because the (actual) numbers will be slightly different.'

When asked about their thoughts on differences between actual and preset values, B2 said:

> 'Things will always be slightly different because…there are other things that we don't measure such as the weather or condition of the equipment on a brewing day that could affect the brewing process…We could try to be exact of course but I don't think that's critical as far as our beer tastes good.'

In these cases, the brew sheets served more as a checklist or guidelines for their existing practices than as a site for systematically and accurately collecting data. By adjusting their targets for interdependent variables instead of writing in the actual values for the preset values, the brewers left many cascading divergences and changes in data values unrecorded, reducing the reliability of the brewing data. They saw these unrecorded changes as a natural, tolerable part of their brewing practices to flexibly work with the high variety of interdependent variables.

Inconsistencies in data collection. Brewers and management highlighted issues with the consistency with which the data at the brewery were collected. It was mentioned, for example, that: 'We are a production brewery, so we are meant to be about consistency' (B4); 'If data such as temperature aren't consistent, then making a change based on data is useless because the temperature will continue to be inconsistent' (HB1); 'For consistency [across brewers in how data are recorded], we might need to introduce some automated data collection methods…but at the end of the day, inconsistencies are introduced by people who do things differently' (M2).

In the early stages of the brewery, HB1 created standard operating procedures (SOPs) for core brewing practices, including instructions for data collection. For example, there was a ‘start of day’ SOP (e.g., safety checks - look for anything out of place such as leaks or strange sounds; turn HLT heating to auto; start recirc pump etc. / fermentation checks - yeast dump (if needed); temperature and pH checks; fill in brew sheets, etc.) and an ‘end of day’ SOP (e.g., temperature checks (BBTs and fermenters); fill HLT and turn heating to timer, etc.). The brewers had also initially developed SOPs for specific recipes (e.g., for a sour beer, ‘recirculate wort through HeatEx until temp is below 48 degC’ (later changed to 40)). The brewery, however, had meanwhile relocated. Its current equipment and materials, as well as beer recipes, had changed after the departure of HB1, yet the SOPs remained largely unchanged, including their instructions on data collection. When asked about their SOPs, B3 commented: ‘We know that things have changed…but we haven’t updated (SOPs). To be honest, we hardly look at them now. Once we think we have gotten used to the settings, we just do things.’

Instead of using the SOPs themselves, we observed the brewers using the brew sheets: ‘The brew sheet shows us the real timeline…because memory is terrible, no one should be remembering the process, you should understand the process and what is to follow next’ (B1). They examined previous brew sheets to modify or confirm recipes for a new batch, carried their brew sheets with them throughout the brewing process, and physically attached the brew sheets to fermenters and bright tanks (tanks for clarifying solids from the fermented beer) to record and monitor the fermentation and packaging processes. Variation in when and how each brewer recorded data (e.g., the duration for each stage of the brewing process differed across brewers or batches as they responded to contingencies (e.g., low water temperature), or the grain auger (conveyor for carrying malt from the storehouse to inside the brewery) was stuck)). Such variations in data collection were seen as inevitable ('our numbers will always be slightly different'), more than something to control through SOPs.

4.4.1.3 Unclear variables (and their relationships) and moving targets

In sections 4.4.1.1 and 4.4.1.2, we illustrated how incompleteness and unreliability in the brewery data arose from simply not collecting data, as well as from collecting data but doing so inaccurately or inconsistently. In this section, we show how, even where brewing data were collected accurately, the data could be unreliable and incomplete as a result of (i) unclear definitions of variables; (ii) unclear relationships among variables; and (iii) the need to change variables or their target values as things changed at the brewery.

Variables not clearly defined. During the data workshops that we held at the brewery, the brewers were asked to walk through how they engaged with data on each variable in their work practices. The brewers often debated the definition of the variables. Some examples were variables for measuring beer volume ('volume transferred,' 'total volume,' 'tank volume,' and 'loss volume') on the packaging log, which were critical outcome variables for the yield efficiency model that we were building. The brewers initially defined 'volume transferred' as the volume of beer transferred from a fermenter to a bright tank; 'total volume' as the volume of beer contained in packaged bottles and kegs; 'tank volume' as the capacity of a bright tank minus 'total volume'; and 'loss volume' as the capacity of a bright tank minus the sum of 'total volume' and 'tank volume.' As they tried to articulate and confirm these definitions, a debate occurred regarding what they meant by 'total volume.'

> B2: Well, we need to decide…I know what matters to us is the volume of packaged beer, but I don't think "total volume" should be calculated with the number of bottles or kegs since some of them will be partially filled or damaged, or beer will be lost during the transfer, and so on.
>
> B1: Can't we get "loss volume" by deducting the packaged volume from "transfer volume"?

B3: Then "total volume" is not the volume in the tank before packaging?

B1: "Total volume" would be the packaged volume, isn't it?

B2: Is that what we are measuring now? We just need to decide.

B3: It shouldn't be confusing. Perhaps we need to use (the variables) 'pre-packaging volume' and 'post-packaging volume' instead.

B1: Theoretically we get "loss volume" by deducting the packaged volume from "transfer volume," but it's not really accurate as we get under-filled bottles (that do not count towards the number of packaged bottles) and beer levels (for these under-filled bottles) are different. So we can't just say that we lost 355ml x 10 because we wasted 10 bottles.

B3: We still have to calculate ("loss volume") somehow. We could instead deduct the post-packaging volume from the pre-packaging volume.

B1: Then what about broken or under-filled bottles?

B2: I don't think we can use the data that we have now. How have we been measuring these?

B1: Maybe we all have been differently…We need to look at…[B1 picks out a packaging log and the brewers start inspecting the numbers].

Upon the inspection of several packaging logs, as illustrated in 4.4.1.1, we confirmed that data on the above variables (i.e., 'total volume,' 'tank volume,' and 'loss volume') as well as other related variables ('filled' (the number of packaged bottles) and 'low' (the number of under-filled bottles)) had been recorded only sporadically. In this example, some of the key variables for predicting yield efficiency were not clearly defined or shared among the brewers, leading to different brewers interpreting and recording (if measured) the same variables differently. This lack of clarity in definitions, coupled with sporadic records, affected both the reliability and completeness of data collected on critical variables.

Everything is connected. In our second workshop, we asked the brewers to explain the relevance of each variable to the problem of yield efficiency for which we were building the ML models. We displayed the variables on the screen in the order of their appearance on the brew sheets. One of the purposes of this workshop was to identify and prioritize, from a domain expert's perspective, a set of key variables to be considered for the ML models. We started with the variables at the top of the brew log. A few minutes in, it became obvious that most variables (from 'brewing date' and 'brewer' to the specs of each ingredient in the recipe) could potentially be relevant to yield. As a way of differentiating variables, we asked the brewers to rate the importance of each variable (from highly relevant to not relevant) based on their knowledge and experience. Most of the variables continued to be described as either highly relevant or relevant. As we moved onto variables related to the mash process (the stage where the malt is mixed with hot water) on the screen, they agreed that most of these variables were highly relevant to yield. Relevance could also be direct and indirect, as seen in their following discussion of the relevance of 'yeast food' (a nutrient mixture to keep the yeast healthy through fermentation):

> B2: Yeast food is not relevant…No, wait, wouldn't that affect the fermentation process? Then that can affect yield.
>
> B1: Yeah, it might not have a direct impact but when yeast ferments it could affect some aspects of yield…So indirectly relevant?
>
> B3: Yeah, but then pretty much everything on these sheets (brew logs, fermentation logs, and packaging logs) would end up being relevant somehow.

The interdependence among variables often made it difficult for the brewers to identify the source of a problem and come up with a possible solution by examining their data. This limited the brewers' ability to take a corrective action during a brewing process. B2 also commented on the actionability of their data after the brewing process was completed, when explaining DO values on a packaging log: 'So an alert for anything over 50 could be useful. But honestly, even if we learn that DO is off, there's nothing particular that we could do about it unless we know the reason.'

The ability to reason about relationships among variables is critical to the brewers' ability to collect reliable and complete data. Such reasoning enables tracing and analyzing the sources of inconsistencies and inaccuracies in the data, as well as deducing missing values or variables based on other existing values and variables. Given the complex interdependence among variables, however, the brewers often had trouble pinpointing relationships between particular variables and their relevance to a given problem, limiting their ability to make their data actionable.

Variables and their targets change over time. Changes in brewing practices or the brewery's material settings often required the brewers to add new variables or modify existing target values on the brew sheets. For example, the brewers had started exploring how they could reuse yeast to save on the costs of their decision to shift from dry yeast (inexpensive and slow-acting, but only available in a small number of strains) to liquid yeast (more expensive, but fast-acting and wide variety). B1 explained: 'Our target is reusing [liquid] yeast three times to compensate for the extra cost…we have no experience in reusing yeast, we don't have any data, we are just experimenting.' B4 showed us a paper packaging log for a red ale where the cells for DO (dissolved oxygen) and CO2 (carbon dioxide) measurements were drawn with a pen to record data for experimentation:

> 'This is a special case - we split one batch with reused yeast and one batch with dry yeast, and I added labels "harvest" for reused yeast and "dry" for dry yeast [on the packaging log] so that we could compare.'

Data on DO and CO2 were collected sporadically over the next month and then discontinued. None of this data got digitized. The paper and online brew sheet templates could not easily accommodate *ad-hoc* additions of variables, resulting in manual annotations (e.g., 'homemade partitions' (B4)) on the brew sheets.

The brewers also at times needed to change the target values of variables to ensure the consistency of their beer against shifting material settings. The brewers had been relying on the

Excel recipe formulas created by HB1. They did not feel confident to flexibly manipulate the targets on Excel, even though they were doubtful of the relevance of the targets to the current brewery settings (the formulas were for batches at different scale, for malt from different suppliers, and for the original brewery facility across town): ‘We presume that there would be problems with the formulas, but we don’t really know how to unpack them’ (B1). The brewers commented that these formulas seemed to be building on data that did not properly reflect the settings and operations of their current brewery. B2 said:

> ‘Things didn’t really match. We’ve been creating new types of beer but, for example, the color (of beer according to these formulas) was supposed to be here [pointing at a color on the European Brewing Convention beer color chart] but the actual beer color tended to diverge from that….Or we were supposed to get 3% ABV with 3kg of barley but ended up getting 2% ABV, stuff like that. If the reference is wrong, then there’s no point in collecting and analyzing data, right?’

The brewers were experimenting with several brewing-related software programs to generate more compatible target values for their variables that they thought would build on ‘more accurate’ data.

The above examples show the need to modify variables and target values to account for changes in the brewers’ practices or the brewery settings over time. Changing theoretical values on the brew sheets could negatively affect the reliability of data, given that the brewers often used theoretical values in place of actual measurements. The difference between theoretical values and actual measurements, however, was hardly recorded (see 4.4.1.2).

4.4.2 Data-related challenges of building the ML model

4.4.2.1 Limits to wrangling unreliable, incomplete brewing data

It is well-known that much of the work of data science and ML does not concern the work of building the ML models *per se*, but rather of 'wrangling' (e.g., cleaning, curating, structuring) initially messy data (Kandel et al. 2012). To prepare for building the ML model for predicting yield efficiency at the brewery, the second and third authors likewise performed a number of tasks to wrangle the data that they initially had been given from the brew sheets. We draw attention to how their wrangling tasks varied in costs, benefits, and feasibility.

Data wrangling tasks too costly or infeasible. Based on his data science expertise, author 3 asked author 2 to help with some data wrangling tasks that would be of high benefit, while imposing minor costs to perform. Such tasks included determining whether a variable was relevant or correcting outliers.

Author 2 could perform these tasks based on his familiarity with the dataset and brewery setting or simply by having a quick look at the data: Author 3 asked in a chat app: 'Just one question about brew no., batch no., date: are these variables potentially predictive [e.g., should they be in the ML models?] or just for documentation? [e.g., are they just background information].' Author 2 responded:

> 'Brew no. and batch number could be predictive… we don't necessarily need brew number since it is already mostly captured by date… there can be up to three batches for each brew… we should expect each batch to be an independent trial, but I think it would be interesting to see if that is not the case, so I would include it as predictive.'

In some cases, however, the third author brought up a desire for more or better data where performing the wrangling tasks would have been too costly, or was simply infeasible because the data had not initially been collected at the brewery.

Author 3 (in the chat app): There are quite a few outliers in mash percent [the percent of liquid remaining in the "mash" stage versus the initial amount], e.g., 8.333(%) for (the batch) 190316 B1 Wit (versus 89% average). Can you correct them? I guess it is just the decimal places, but it is better that you double check.

Author 2: I corrected the mash percent outliers [by checking if the decimal points were off, or asking the brewers]; if there were no values, I left it blank - this is the missing data for mash percent.

Author 3: Awesome, no more outliers in mash_percent. but 30 missing values unfortunately.

In other cases, the second author informed the third author about variables that they should keep if at all possible, even if the data were not so reliable or complete:

Author 3: *Can you have a look at the last four variables ("mash_start_temp," "cool_temp," "post_cooling_ph," and "final_ph") at the bottom [of the report of the initial ML model results] if we could discard them? They have >30% missing values*. Author 2: *We can take out cool_temp and final_ph, especially since post_cooling_ph is pretty close in the brewing process… post_cooling_ph is the one I'd really want to keep [since the brewers emphasized the importance of pH]… if possible, it would be good to keep mash_start_temp since the (outcome) variable (for the ML models) is mash percentage.*

Author 3: OK, will impute post_cooling_ph since it has 59.64% NAs [missing values].

During a call to discuss the results of the first sprint, author 3 asked whether the values of two critical features, gravity (a measure of the density of the beer used to calculate the alcohol percent (ABV)) and pH, could be trusted. Author 2 remembered B4 mentioning that the procedures for both of these features were not standardized at the brewery. B4 also suspected some of the data may have been imputed, rather than actually measured. Author 2 and author 3 decided that getting trustworthy data on these features would be infeasible. They discussed having conversations about the values with the brewers, but concluded that such conversations would not do much good without an overhaul of the work practices of the brewery for standardizing procedures for measuring and collecting data on these two variables.

Noting that the initial data had a large number of features (e.g., 85 columns of data about brewing variables) relative to the number of observations (289 rows for the brew log, corresponding to the number of batches included in the dataset), author 3 pleaded for even a small number of additional rows of data to improve the model predictions: 'Could you check whether you could guess the missing brewers [the brewer's name was not entered in five of the brew logs] and yeasts. I wouldn't ask but I just try to get as much data as possible.' Author 2 had gathered the brew sheets six months earlier. He knew the brewery would have six months' more batches worth of data, or probably around 50 batches. Yet collecting more data would take weeks since it was in physical binders and had to be manually entered and cleaned. He decided that the benefits to the model predictions (e.g., more rows of data would lead to more robust predictions) would be outweighed by the data collection costs.

In his notes analyzing the results of the second sprint, author 2 expressed concern that the data in the brewery were currently collected only on paper brew sheets, such that it could only be updated by further manual entry into the Google Sheet that they had been using for building the ML model. He suggested syncing their data from the Google Sheet to an SQL database, where they could add user-friendly data entry tools for the brewers. Author 2's concern was that, even if the sprints led to good ML model predictions, the predictions would quickly deteriorate if the data were not updated. Author 3 worried that adding an SQL database would both be costly to build, and increase data processing costs for running the ML model: 'As a (former) IT project manager, I was accountable for all costs, every penny of it… and [SQL databases] are awfully slow… [it] would be a bit of overkill…. I need the data in one big table, that's it.' They decided to keep the brewing data in the original Google Sheet.

In the remainder of section 4.4, we examine the challenges that the unreliable, incomplete brewing data posed for author 2 and author 3 in building the ML model of yield efficiency at the brewery.

4.4.2.2 Challenges for the domain expert to inform the data scientist

In this section, we describe how author 2 (the proxy 'domain expert'), faced challenges in informing author 3 (the proxy 'data scientist') about individual features of the brewing data as they built ML models of yield efficiency at the brewery. First, we establish that informing author 3 about features of the ML model depended on author 1 recognizing the need to interpret complex nonlinear relationships in the data. Second, we show how author 2's ability to interpret nonlinearities was complicated by the lack of reliable and complete brewing data.

Establishing nonlinearity in the brewery's yield efficiency model. After author 2 and author 3 wrangled the brewing data to the extent that they could (with the data still unreliable and incomplete), they began a series of eight 'sprints' to build an ML model of the yield efficiency problem at the brewery. For each sprint, author 3 ran five models (a baseline linear regression (non-ML) and four ML models) on different sets of features that author 2 and author 3 jointly selected for each sprint. In a conversation after the final sprint, author 3 emphasized to author 2 that ML models are distinct from linear regression in that they capture nonlinear relationships in data. Author 3 noted that, in all eight sprints, 'The [four] machine learning models easily outperformed the [baseline] linear model,' meaning that all of the ML models' predictions of brewery yield efficiency had higher accuracy (e.g., explained more of the variation in yield across batches) than the simpler linear regression model. Since the ML models easily beat the baseline linear model, he concluded that 'there is strong evidence that the features we selected from the dataset for all sprints have a nonlinear relationship to [yield efficiency].'

Evaluating nonlinearities in unreliable, incomplete data using XAI. Challenges for author 2 arose in that the XAI results for the same features varied wildly (i) across the five different XAI techniques; (ii) across the sprints, and (iii) depending on which of the four ML models the XAI techniques were applied to.

In one example, after the third sprint, author 2 looked at an XAI visualization of the importance of the feature 'IBU' (a measure of the bitterness of a beer recipe) for each of the four ML models. He noted that 'IBU' was the most important feature for the 'GBM' (gradient boosting machine) model, but of almost no importance for the 'RF' (random forest) model. Author 2 posted on the sprint report:

> 'The GBM vs. RF differs a lot — so how do I weigh these and give a recommendation?... This will be very confusing for the brewers... for most of the other [XAI results] I also have no idea. They seem totally different from the GBM, so I have not much basis for interpreting what I'm looking at.'

Author 3 responded in the sprint report by explaining that: 'XAI (techniques are) not a silver bullet for simplicity. Correlations react sensitively to changing data distributions. Linear correlations are always the same… but each of the ML models had much different nonlinear data distributions.' He added to the same document that the incompleteness of the brewing data was a driving factor in making the complexities of the data distribution extremely hard to interpret:

> 'The missing values and questionable data impose a really strong reliability issue… you emphasized you really want to keep certain [features], which I did. But you must be aware that such decisions have direct consequences on the reliability of results given the data we have. So you will naturally see results changing dramatically from sprint to sprint.'

Author 2 responded in the document that 'I would recommend we talk through our overall approach in person before going through another round of evaluating [which] features [to keep in the ML model]'. The following week, author 3 then tutored author 2 for several hours on using the details of each of the XAI visualizations, going over the relationship between the data distributions specific to each of the four ML models, and how knowledge of these distributions could be used to interpret the visualizations even if they changed. Author 2 stated 'now that I have a better understanding of the XAI visualizations (of nonlinearities in the data),' he could use his domain knowledge to recommend adding back in three features (IBU, Target Original Gravity, Brewer) he had removed after sprint 1 due to poor linear correlation with yield.

4.4.2.3 Challenges for the data scientist to make the ML model predictions actionable

In this section, we describe how author 3 was able to generate accurate predictions in the ML model of yield efficiency by the end of their eight 'sprints', yet the predictions were perceived by the brewers as not actionable (e.g., for taking corrective actions in their brewing practices to improve yield). We then connect this gap between the ability of the data scientist (author 3) to generate predictions with the lack of actionability for the domain experts at the brewery to the unreliable, incomplete brewing data.

ML model predictions were accurate. Despite the challenges for author 2 in informing author 3 about the features of the ML models, their eight sprints led author 3 to generate progressively more accurate ML models. After the fourth sprint, with the set of features reduced from 47 initially to 14, an R-value (a measure of correlation) between 45% and 53% across the four ML models. Author 3 remarked in the sprint report that this was 'pretty good [accuracy] for so few predictors, but another sign of messy measurement [in the brewing data]. We can make [the model] better by removing even more features to further reduce the noise.'

By the eighth sprint, author 2 and author 3 had settled on a set of eight features out of the initial 47 for the final set of ML models to be used to predict yield efficiency at the brewery, with the R values still remaining similar to the fourth sprint despite the parsimonious model. Author 3 excitedly remarked to author 2 on chat: 'You can be happy… we got an R [a measure of correlation] over .50 [in the GBM] which is very astonishing to me, given the poor data quality… the improvement for [the] GBM [model] is quite impressive.'

ML model predictions were not actionable. The following week, author 2 and author 3 held a workshop on Zoom with all four brewers. They first presented the results of the ML models, explaining that the predictions were accurate in a statistical sense, and could be further improved with more data. They then asked for the brewers' general reflections on the ML models, and began a discussion to explore specific ways to make the predictions actionable for their brewing practices.

All four brewers responded that they needed much more than a single yield prediction for a batch to be able to take corrective actions. The key issue was that, for the brewers, solving the yield efficiency depended on being able to predict yield issues at many different points of the production process:

> B3: [We] really [don't] know what variables need to be changed or what actions could be taken during the process (to improve yield), so we just wait until the process [batch] is completed.
> B1: We may "feel" that something is wrong [during a batch], but right now we only use the outcome [the final beer that is produced] to evaluate the [brewing] process…which makes it too late to do anything. But if we can [use the ML models to] get real-time indicators of something going wrong across different stages, that would be helpful.
> A1: So, if something goes out of range at a particular brewing stage according to the ML model, then this could trigger a suggestion to take some action specific to that stage?
> B1: Yes, that would be really helpful. Say we get a lower pH, but we're not sure if we need to add more acid or how much to keep yield [normal]. So if the machine can generate data and make predictions, it could suggest how I could correct [the acid] on the fly. We are not really documenting this kind of variance - nor are we making any corrective actions at the moment.

ML model predictions could not be made actionable due to lack of data. Based on the workshop, author 2 and author 3 concluded that making the ML model predictions actionable depended on extending the initial ML models to a much larger set of outcome variables that corresponded to the many different stages of the brewing process. In the sprints, author 2 and author 3 had used 'mash percent', which provided a yield efficiency measure at a critical stage of the brewing process, but was only one such stage. In an online video chat the day after the workshop, author 2 communicated to author 3 that he had pushed to focus their sprints on "mash percent", because it had fewer missing values than all other outcome variables in the brewing, even though it still had around 30% missing values. Author 3 remarked that we would 'rather have fewer observations (of outcome variables) but without data entry errors, especially of the target variable [mash percent]…A few of these errors, bad data, and the [outcome variable] cannot really be used'.

Another issue concerned evaluating which features should be in the ML models. Author 3 pointed out that the best way to evaluate each variable of the brewing data would be to run many different models, both with and without the variable, and compare the results. For efficiency reasons, however, it was only possible to make a small number of changes to the features included in the model relative to the total possible permutations of the 47 initial variables. The issue was that, while author 2 informed author 3 regarding the definitions or relevance of individual features, author 2 could not easily communicate the logical relationships among features in the brewing practices that would have helped simplify the choices of which models to run. As author 3 remarked:

> 'What would have been [more efficient] is a schematic diagram that mapped the features to their relation to the brewing process and a very simple description of the brewing process, which makes

clear which variables should be more important for the outcome. Then every sprint, you [author 2] have to explain it back to me.'

Author 2 and author 3 concluded that, given the brewers' feedback from the workshop, what the brewers wanted from the ML models would require more reliable and complete data, including documentation of the logical relationships among features of the data.

In this section, we looked at how the second and third authors faced data-related challenges to building an actionable model for the brewers. We began by establishing how it was too costly or infeasible for author 2 and author 3 to make the brewing data reliable or complete after it had been collected at the brewery. We then described two ways in which unreliable, incomplete data complicated the two-way interaction by author 2 and author 3 in building the ML model. First, for author 2 (playing the role of 'domain expert'), lack of reliable, complete data meant that his ability to select features for an actionable ML model depended on knowing details about statistical distributions that were too technical for him. Second, for author 3 (playing the role of 'data scientist'), lack of reliable, complete data led him to generate model predictions that were accurate in a statistical sense, but not actionable for the brewers. Finally, we described the brewers' reflections on why they did not find the predictions of the ML model to be actionable in their brewing practices.

Lack of reliable, complete data on target ranges led author 3 to generate model predictions that were accurate in a statistical sense, but not actionable for the brewers. The challenges of two-way collaboration (between data scientists and domain experts) in selecting data for, and generating predictions with, ML models point to a need for understanding why the brewery lacked reliable, complete data in the first place.

4.5 Discussion

4.5.1 Towards sustainable data science: domain experts as 'owners of data'

Much of the practitioner discourse on data science related to sustainability invokes industrial metaphors of efficiency. Data flow through carefully engineered 'pipelines' (e.g., Polyzotis et al. 2017; Amershi et al. 2019b). The efficient flow of data, in this view, is sustained by a virtuous cycle in which data and algorithms are mutually reinforcing sources of feedback. At the center of the work of sustaining this virtuous cycle between data and algorithms has so far been the data scientist, cast variously as coder, wrangler, or modeler, and informed by the domain expert.

In contrast, CSCW and HCI researchers have long emphasized that the contingent nature of domain experts' work practices and settings confounds plans to engineer technologies for sustainable use (Fischer et al. 2009; Segal 2009; Suchman 1983). Domain experts, as end-users, are the ultimate 'owners of problems' (Fischer et al. 2004, p.35; Mørch and Mehandjiev 2000; Segal 2009). The situated knowledge and experience that they bring is essential to the sustainability of technological tools. This emphasis on domain experts as owners of problems has continued in emerging research on data science in CSCW and HCI (e.g., Gil et al. 2019; Amershi et al. 2014).

We propose that the design of sustainable data science tools be guided by a view of domain experts not only as 'owners of problems' to be solved (e.g., by ML model predictions), but as 'owners of data' to be used. Domain experts 'own' data in that they do not only inform data scientists about what data mean or use predictions from algorithms. They also engage in data practices — the situated work by which domain experts create, collect, manage, make sense of and otherwise use data in their ongoing work practices and settings (e.g., Zhang et al. 2020; Muller et al. 2021). While the accuracy of predictive algorithms depends on reliable, complete datasets, the reliability and completeness of a dataset depends on domain experts' data practices (Amershi et al. 2019b; Sambasivan et al. 2021). It follows that the design of data science tools can benefit from aligning the data practices of domain experts, as 'owners of data', with data science activities. By shifting the central role from data scientists to domain experts, the virtuous cycle that sustains data science activities expands to include how their data practices mediate the interplay between data and algorithms.

Drawing on ethnographic fieldwork at a craft brewery, an 'owners of data' view directed us to three situations (and their contributing factors) where the brewers' data practices led to unreliable, incomplete data: (i) variables defined but data not collected (due to lack of resources, and to lack of prioritization or integration into work practices); (ii) data on variables collected but not reliable (due to use of preset values in place of measurements; divergence of preset values from actual measurements, and inconsistencies in data collection); and (iii) unclear and changing variables (due to unclear definitions of, or logical relations among, variables, and to changes in the brewers' material work practices and settings). Drawing on a pilot ML project at the brewery, our study examined the effects of the brewers' data practices on data science activities, finding that unreliable, incomplete brewing data: (i) could not be adequately wrangled after the fact, leading to challenges (ii) for the 'domain expert' to make sense of the ML model predictions; and (iii) for the data scientist to make the ML model predictions actionable.

Thus far, support related to domain experts' data practices largely concerns tools (e.g., interactive visualizations) for tasks such as wrangling (e.g., discovering or cleaning data) or informing data scientists. These tasks are mostly performed after a dataset has been provided by domain experts or created elsewhere. *Post-hoc* support may be inadequate as a basis for sustainability if the data in datasets are unreliable or incomplete (e.g., Pine and Liboiron 2015).

By tracing the sources of unreliability and incompleteness of a dataset to domain experts' data practices, our study contributes insight into challenges to sustainability in data science activities. In addition, we contribute design implications for sustainably aligning domain experts' data practices with data science activities.

4.5.2 Challenges for sustainability in data science activities

4.5.2.1 Misalignment between logical and statistical uses of data

Our findings show that the brewers used data in the brew sheets mostly by referring to logical relations among small numbers (2-3) of variables. For example, they made sense of actual versus theoretical values by referring to the values of a few related variables. In the data workshops, the importance of a variable was explained by how it related sequentially to a few other variables (e.g., X affects Y in the mash process; X + Y is roughly equal to Z when beer is transferred from fermenter to bright tank). The brewers, however, found the logical use of data challenging. Variables were related in multiple, causally ambiguous ways (e.g., for the yield efficiency problem, 'pretty much everything…would end up being relevant somehow'; 'there's nothing particular that we could do about [yield efficiency] since we don't know the reason'). Certain variables were not clearly defined (e.g., what was meant by 'total volume'), causing confusion over how these variables needed to be calculated or measured. Modifications to out-of-range variables during a batch were done mostly by ballparked, on-the-fly calculations that went unrecorded, as seen in the case of modifying 'pre-boil volume'.

In the logical use of data by the brewers, we observed a misalignment with the more statistical use of data in the pilot ML project. Variables in the ML models were evaluated primarily for their contribution to predictive accuracy (e.g., author 3: 'Are these variables potentially predictive?', 'We can make [the models] better by removing even more variables to further reduce the noise'). In our workshops with the brewers, this statistical use of data did not resonate, as it was not clear how even precise predictions could be mapped to corrective actions. Though the variables selected for the ML models were informed by author 2's domain knowledge (which in turn had been informed by the brewers), author 3 could only run a finite number of models. To select models with outputs that were considered more actionable by the brewers, author 3 expressed he would have needed to understand how variables were logically related with respect to the brewers' practices and settings.

Design implications: supporting different uses of data. Design support for domain experts in the prior research on data science has tended to assume a given dataset. The focus is on when and how to bring in the domain expert to enable the data scientist to make sense of (abstract) data. Our focus on domain experts' data practices highlights the need for design also to address misalignment between the logical use of (situated) data by domain experts and the use of data for making statistical predictions by data scientists. To design support for aligning these different uses of data, we propose enriching the ability of domain experts to construct and refer to logical relations among variables in data science tools.

Support could include (i) interfaces for defining, manipulating, and sharing visualizations of logical relations among variables that account for domain experts' work practices and settings (i.e., use of schematics of work practices and settings) or (ii) database features for domain experts to explore variables in logical terms (i.e., querying tools). Such tools may enable more effective and sustainable two-way interactions between domain experts and data scientists. If domain experts find more ways to logically explore their data, the benefit would be richer ways of informing data scientists. Data scientists who are able to identify logical relations relevant to practice might transition from optimizing predictive accuracy to selecting variables or models that give rougher but more actionable predictions.

4.5.2.2 The value of reliable, complete data may be unclear to domain experts

In building the ML models, the 'domain expert' (author 2) was unable to interpret visualizations of the importance of variables based on supposedly 'explainable' (XAI) techniques. The XAI visualizations could not be interpreted based on domain knowledge since they varied wildly with even small changes to the set of variables or type of ML model. The wild variations were rooted in how ML models capture nonlinear relationships that are highly sensitive to the reliability and completeness of the data. For a domain expert to interpret the importance of variables in the brewing data, more and better data would have been needed to stabilize the XAI visualizations across the 'sprints' and types of ML models.

The brewers we observed, however, valued data primarily as reference points to guide their brewing practices (e.g., Is there anything in the brew log for this batch that seems off?). They did not see the value of collecting data that would be reliable and complete from a data science perspective (e.g., for building the ML models of yield efficiency). Without answers to questions of 'what can I do with this ML prediction', the brewers had little incentive to collect data for ML models beyond the variables that they used as guidance during their brewing practices (but which were needed, for example, for modeling yield efficiency).

The brewers' tolerance for incomplete data was evidenced by how they would (i) defer the repair or replacement of broken measuring instruments; (ii) work around not having equipment or resources to measure variables by relying on experience or proxy indicators (e.g., estimating 'volume transferred' based on past batches or on how much water was added to top up a bright tank); and (iii) easily fall into not collecting data on related variables in the event of disruptions (e.g., a broken flow meter). The brewers' tolerance for unreliable data collection was evidenced by (i) their frequent use of theoretical values in place of actual measurements; (ii) how in some cases they relied on values calculated in their heads based on experience ('we can kind of tell if the [DO, TPO values] are off by looking at the amount of foam that forms during bottling'); and (iii) the lack of clearly defined SOPs for data collection, despite vocalizing that this resulted in inconsistencies in data collection. In contrast, data collection on variables (e.g., pH, plato) perceived as vital to monitoring the brewing process tended to be more reliable and complete for each stage of the brewing process, even though measuring them required effort.

The implication is that, to induce a more virtuous cycle between data practices and data science tools, design support is needed for domain experts to see the value of data science tools, and in particular how this value is affected by the reliability and completeness of the data they collect.

Design implications: seeing the value of data science tools (to see the value of data). A central theme in CSCW, HCI and STS research on data science has been concerns (e.g., bias, fairness) with the lack of interpretability of models or algorithms (Rudin 2019). Extant research, however, has mostly studied interpretability in contexts where a single prediction is made (e.g., a credit scoring algorithm that predicts who to approve for a loan). Less has been said about the interpretability of data science tools in our study's context of domain experts and their situated, ongoing work practices. While the ML models in our study similarly lacked interpretability, the biggest concern for the brewers was not bias or fairness *per se*, but the unclear value of the ML model predictions to their brewing practices.

Going beyond making predictions, data science tools may enable the brewers to formulate a range of problems that they face in their situated work practices and settings in terms of relevant variables of data (see Passi and Barocas 2019). For instance, templates, examples, or use cases could point to where the statistical predictions of ML models can be helpful, as well as other outputs such as visualizations (e.g., fermentation trends) or queries (e.g., comparing pH values across multiple batches of a particular recipe). Such tools would make it possible for domain experts to build connections between problems and data using data science tools, while keeping them at arm's length from the nonlinearities in data science. Second, the tools could visualize how the value of the data science tools varies with the reliability and completeness of data to help domain experts see the relationship between their data practices and the effectiveness of data science activities. The idea would be to induce a virtuous cycle by enabling domain experts to see the value of data science tools to their problems, and the value of data to their use of the tools.

4.5.2.3 Managing data as they are collected (not wrangling them after)

Even where the value of reliable, complete data was clear to the brewers, the brewers lacked data management tools to flexibly adapt the brew sheets to changes in their brewing practices (e.g., shifting from dry yeast to liquid yeast) or material settings (e.g., weather, grain supplies, equipment). For example, when the brewers experimented with data on new variables, they found it cumbersome to modify the brew sheet template files, instead adding the variables on the margins of the paper brew sheets in an ad hoc fashion. Since the values were only recorded sporadically, the new variables could not be used in the ML models. Another example concerned discrepancies between the actual and target values (e.g., for the 'final gravity' of a beer) of variables. When the brewers did not know what contributed to these discrepancies, modifying the target values became challenging or was simply not done (e.g., see section 4.4.1.3). These targets would have been key outcomes to predict in the ML models, but were not usable since the target values were absent or not reliable.

As noted in section 4.2.5, recent research in CSCW and HCI increasingly emphasizes the importance of contingent, domain-specific data work in data science activities (Amershi et al. 2019b; Bopp et al. 2017, etc.). Data management tools for data scientists (e.g., workflows for data 'pipelines' or ML 'lifecycles'), however, remain oriented towards 'one-off applications' (Polyzotis et al. 2017) in which data are collected in a 'largely stable world' (Marcus 2018). These tools reflect a view of data-related bottlenecks to sustainability as mostly concerning challenges of modifying datasets (e.g., wrangling) after the data are collected and enter the 'pipeline'. In our setting, the biggest challenge was not how to efficiently wrangle given data, but how ongoing changes to the variables or values of the brewing data impaired the reliable, complete collection of data.

Design implications: data management tools for domain experts. To enable data to be modified sustainably amid ongoing changes in practices and settings, data science tools could support domain experts to have a central role in data management. First, data entry interfaces tailored to work practices (e.g., specific to each stage of the brewing process) could reduce the burden for, and increase the reliability and completeness of, data collection by domain experts. Such interfaces could promote sustainability in data science activities if more reliable, complete data collection reduces the need for *post-hoc* data wrangling. Second, tools could enrich support for domain experts to add, remove, or modify variables in templates for recording data (e.g., the brewers' recipe templates). Finally, these user-level tools need to be linked within an integrated database management system that syncs any changes to individual templates or records while maintaining integrity of the data.

While the idea of a database management system has been around for decades, such systems have typically been implemented at the level of an organization (e.g., an ERP system) or large technical endeavor (e.g., data feeds of large technology firms, or databases for basic science), rather than at the level of domain experts' data practices. By designing tools specific to data management by domain experts, the idea would be to manage data in a way that can account for changes in practices and settings, thereby promoting sustainability in data science activities.

4.6 Limitations and future research

Our study built on field data from a single craft brewery site in Korea to examine how a group of craft brewers, as domain experts, collected and used data in their practices and settings. All our brewers were male in their 30s and 40s, with two of them being non-Korean. These sociocultural particularities may pose limits to generalizing our findings and analyses to large companies, research organizations, or technology firms where their work is more easily standardized or data collection and management systems are highly centralized or specialized. We believe, however, that the theoretical framework of domain experts as owners of data would be a useful lens for studying the relationship between data practices and data science activities, and informing the design of sustainable data management or data science tools for a range of sites where domain experts themselves collect and work with their data in their practices and settings. Future research may draw on our framework to study other types of domain experts to identify differences and similarities in their data practices and considerations for data science activities.

We acknowledge that, while we developed ML models based on the brewing data as a pilot project, these models were not implemented or deployed at our site. These models were used to explore and identify opportunities and challenges regarding the possible introduction of data science to the brewery. Although the brewers were briefed and provided their feedback on the processes of building the ML models and the results of the models, they did not directly engage with them. We also acknowledge that tools and technologies for ML and XAI are changing rapidly

(Arrieta et al. 2020). It is possible that our insights regarding the process of building ML models are specific to the current state of the art. These limitations present opportunities for future research on developing, testing, and evaluating a range of data science and ML tools to study domain experts' data practices and how these practices mediate data science activities.

4.7 Conclusion

This chapter proposed a view of domain experts as owners of data in data science systems. This view makes central the role of domain experts' data practices for sustainable data science activities. Drawing on fieldwork at a craft brewery in Korea, the findings identified situations where brewers' data practices led to unreliable, incomplete data. The study examined how such data mediated data science activities by drawing on a pilot project to build ML models at the brewery. The analysis of the project showed how unreliable, incomplete brewing data led to challenges for wrangling data, understanding ML model predictions, and making the predictions actionable. Based on these findings, the chapter discussed three challenges for sustainability in data science activities from a domain expert-driven point of view and provided design considerations to address these challenges. By proposing that data science activities be driven by domain experts as owners of data, the study aims to inform the design of data science tools that domain experts can use in a sustainable way, a central objective of research in CSCW and HCI.

5 Expectations of People with Visual Impairments for Image Descriptions

In the previous two chapters, I explored how craft brewers, as a particular type of data users or potential end-users of data science systems, worked with data in their practices and settings. I focused on how their data practices mediated the effectiveness and sustainability of data science activities.

This chapter[3] turns to a setting where data users have little or no access to the properties of artifacts of interest - people with visual impairments (PVI) accessing photos on their personal devices through image descriptions. How do these humans use data to access, make sense of, and take actions regarding their artifacts that are highly opaque? How should AI-generated image descriptions be designed and provided to help them meaningfully engage with their artifacts?

Thus far, research has studied what PVI expect in these descriptions mostly regarding functional purposes (e.g., identifying an object) and when engaging with online, publicly available images. Extending this research, I interviewed 30 PVI to understand their expectations for image descriptions when viewing, taking, searching, and reminiscing with personal photos on their own devices. The study shows how their expectations varied across photo activities and often went well beyond identifying objects in photos. Based on the findings, I propose design opportunities for generating and providing image descriptions for personal photo use by PVI. The design

[3]This chapter has been published as Jung et al. (2022b). I proposed and conducted the interview study, with the help of Junbeom Kim in recruiting participants. I analyzed the findings and wrote the manuscript with feedback from the co-authors (Tom Steinberger, Junbeom Kim, and Mark S. Ackerman).

opportunities for PVI also point to novel support for the sighted for using image descriptions to enrich their experience of photos.

5.1 Introduction

People with visual impairments (PVI) access images through descriptions. Image descriptions may refer to any spoken or written account of photos generated by humans, computers, or a hybrid of the two (Bigham et al. 2010; Zhong et al. 2015; Simons et al. 2020; Fang et al. 2015; Wu et al. 2017; Stangl et al. 2018; Salisbury et al. 2017; Morris et al. 2018; Guinness et al. 2018), but the term is most often used interchangeably with "image caption" or "alt text" to refer to computer-generated text about an online image (Stangl et al. 2020; Stangl et al. 2021). A number of tools for computer-generated image descriptions of photos have recently been developed amid the proliferation of smartphones and social media platforms. This trend has led to an important research agenda in HCI and accessibility studies of how to design image descriptions that support PVI to engage with photos across various settings and contexts (Jayant et al. 2011; Harada et al. 2013; Morris et al. 2016; Morris et al. 2018; Gurari et al. 2020; Antol et al. 2015; Bhattacharya et al. 2019; Zeng et al. 2020).

Thus far, research has mostly examined image descriptions for PVI regarding online, publicly available images, and where the purpose is mostly functional (e.g., to identify and recognize visual elements such as an object or a person). Extending this research, we interviewed 30 PVI to understand their expectations for image descriptions when viewing, taking, searching, and reminiscing with personal photos on their own devices. Developing insight into personal photo use of PVI is important for designing image descriptions, as a range of smartphone image recognition applications and accessibility features are becoming available for PVI. These tools are being designed to help PVI engage with the content of existing photos or the camera frame, providing audio descriptions of textual information, products, people, scene, colors, and light (e.g., Seeing AI; TapTapSee; Sullivan+) as well as metadata about photos such as the date and location that they were taken (e.g., VoiceOver Recognition; TalkBack).

In our findings, we show how expectations for image descriptions in our participants' photo activities often went well beyond identifying objects in photos. Participants wanted image descriptions to help them reflect on, imagine, and share potentially multisensory experiences triggered by the photos' content. We explain how and why these expectations also varied across photo activities.

Based on our findings, we propose design opportunities for each photo activity (viewing, taking, sharing, and reminiscing with photos) and three broader design considerations to generate and provide image descriptions for personal photo use by PVI. We discuss how these design opportunities and considerations for PVI, by going beyond functional goals of recognizing what is visible in the photo, may also point to novel support for the sighted to use image descriptions to enrich their experience of photos.

5.2 Background

In this section, we examine prior literature on image descriptions, support for photo use by PVI, and the use of mental images by PVI to engage with visual information.

5.2.1 Image descriptions for PVI

Much research on accessibility and computer vision explores tools to identify and describe visual elements of images to support engagement with photos by PVI. Tools include: a personal object detector that trained on photos of objects taken by PVI themselves (Kacorri et al. 2017); a system called RegionSpeak that stitches together multiple images to offer descriptions of spatial relationships between objects (Zhong et al. 2015); an attention-based neural image caption generator that captures salient visual information (Xu et al. 2015); and the Accessibility Bot, which identifies and describes the identity and facial expressions of the user's social media friends included in images (Zhao et al. 2018). Image descriptions can be created manually by humans through crowdsourcing or volunteering (e.g., Bigham et al. 2010; Brady 2015; Zhong et al. 2015; Simons et al. 2020; Zeng et al. 2020; Lee et al. 2020), automatically by AI models trained to capture particular aspects of images (e.g., Fang et al. 2015; Wu et al. 2017, Stangl et al. 2018; Guinness et al. 2018; Wang et al. 2021), or by combining AI-generated descriptions with human inputs (e.g., Morris et al. 2018; Salisbury et al. 2017; Yuksel et al. 2020).

Prior studies in accessibility and HCI have identified a wide range of content that PVI want included in image descriptions, ranging from salient objects, people (e.g., identity, age, gender, clothing, facial expression), settings, actions, atmosphere, colors, and photo quality (e.g., lighting, composition) to the spatial relationships among objects and how people are interacting (Petrie et al. 2005; Ramnath et al. 2014; Zhong et al. 2015; Morris et al. 2016; Voykinska et al. 2016; Wu et al. 2016; Wu et al. 2017; Zhao et al. 2017; Stangl et al. 2020). Studies have found that PVI want to use descriptions not only to identify and learn about elements featured in images, but to evaluate the quality or meaning of images, organize and manage photos, and to participate in social interactions (Bigham et al. 2010; Brady et al. 2013; Harada et al. 2013; Wu et al. 2014; Voykinska et al. 2016; Zhao et al. 2017; Stangl et al. 2018; Bennett et al. 2018b; Gurari et al. 2020).

Extant research also found that how the content of descriptions is structured and presented affects the experiences of PVI (Gurari et al. 2020). Recommendations include ordering words according to their relative importance in a given context (Petrie et al. 2005), providing descriptions in complete sentences rather than a list of words (Wu et al. 2017), and structuring descriptions consistently at varying levels of detail (Stangl et al. 2018). Others suggest that spatial information could be provided clockwise (Kim et al. 2017; Sáenz and Jaime Sánchez 2009), or localized to allow for the user to selectively and interactively explore the content of images (Jing et al. 2015; Zhong et al. 2015; Seeing AI). Morris et al. (2018) present a "taxonomy of properties" to be considered when constructing and delivering descriptions, which includes interactivity, stability, and personalization.

Effectively generating and providing image descriptions for PVI, however, faces several fundamental challenges. The first challenge concerns the "*description gap*" (Stangl et al. 2020, p.1). What people want or need in descriptions varies widely depending on the context of image use, visual conditions, or personal preferences or backgrounds (Petrie et al. 2005; Voykinska et al. 2016; Zhao et al. 2017; Stangl et al. 2020; Gurari et al. 2020). Descriptions of an image tend to be partial, selective, or subjective, making the validity and accuracy of descriptions often difficult to evaluate (He and Deng 2017). Descriptions of certain elements in images, such as people's identity, may need to change over time for them to stay accurate or relevant (Wu et al. 2017). To address this description gap, studies have begun to explore possibilities for generating context-dependent, personalized descriptions (Stangl et al. 2021; Gan et al. 2017; Mathews et al. 2016).

Another challenge concerns the appropriateness of descriptions. Studies point to the uncertainty and lack of agreed-upon guidance in describing potentially offensive or sensitive content (Wu et al. 2014; Simons et al. 2020; Stangl et al. 2020), or content that could cause privacy concerns (Cosley et al. 2012; Morris et al. 2016). Bennett et al. (2021) emphasize the negotiated nature of describing the appearance of potentially marginalized populations. They suggest PVI should be given control over how they themselves may be described and who would be given access to different descriptions. Hanley et al. (2021) compare policies for alt text generation and show how automated and manual approaches differ in regard to identity-related content. Finally, Edwards et al. (2021) used fictional characters, each of whom had multiple types of disabilities, to evaluate descriptions that could more properly capture the characters' complex identities.

Taken together, extant literature examines the visual elements in an image that might be identified, and how these elements could be described and delivered. This literature, however, has mostly studied how PVI want to engage with publicly available images online. This literature also tends to focus on the use of descriptions for recognizing the content or evaluating images to decide how they could be shared or otherwise used for specific tasks. What remains under-explored is the role of descriptions in the engagement with personal photos by PVI on their own devices (either taken by them or shared with them), which may go beyond recognizing or evaluating the content to include use such as reminiscing or otherwise subjectively experiencing photos.

5.2.2 Photo support for PVI

PVI want to engage with photos for various reasons, from recording and sharing memorable events to capturing information that they cannot easily access (Jayant et al. 2011; Bennett et al. 2018b). Several studies have developed photo support tools for PVI. Tools to support photo taking provide, for example, audio or haptic feedback on the position and size of faces in the frame to help PVI aim or move the camera and compose the content in the frame as intended (Jing et al. 2015; Balata et al. 2015; Vázquez and Steinfeld 2012). More recently, Lim et al. (2019) developed a system called TouchPhoto, which provides additional features of embedding audio tags in photos and displaying salient facial features using electrovibrations. Iwamura et al. (2020) proposed VisPhoto, which takes an omni-directional image from which a selected region can be cropped into a photo *post hoc*.

Studies have also explored support for organizing and browsing personal photos. Harada et al. (2013) developed a system that captures photos with audio recordings of ambient sounds and memos to help PVI users browse through their photos. Using a similar photo-audio system called VizSnap, Adams et al. (2016) found that ambient sounds and voice memos were some of the most helpful cues for participants in organizing and browsing through their photos.

Given their practical goal of making photos accessible to PVI, extant research on photo support for PVI tends to focus on creating and providing additional data (e.g., audio, metadata) to access photos and perform particular photo activities. Less studied, however, is how PVI themselves experience photos through such data. For example, while extant research considers audio descriptions as a key element of designing systems to support photo activities by PVI, little is known about how PVI actually experience these descriptions.

One example that points to the importance of how PVI might experience photos is the HCI research on reminiscing with photos. Reminiscing is "the process of recalling personally experienced events from one's past" (Webster and McCall 1999, p.73) motivated or triggered by "emotional or sentimental reasons" (Sellen and Whittaker 2010, p.73). Photos, in their conventional association with memory (e.g., Sontag 1977; Barthes 1981), have been considered as a key resource for reminiscing. Much research stresses the positive role of reminiscing activities in helping us explore, understand, and maintain our identity over time (Cosley et al. 2012; Webster and McCall 1999), and in providing a sense of control over our lives (Heckhausen and Schulz 1995).

As people take, keep, and "forget" about an increasing number of photos amidst the proliferation of photo taking and sharing tools, researchers in HCI and interaction design have attended to the ability to revisit or rediscover photos as an important condition for reminiscing with them (Petrelli and Whittaker 2010; Whittaker et al. 2010; Sarvas and Frohlich 2011; Lindley 2012; Frohlich et al. 2013; Odom et al. 2014; Van House 2016; McGookin 2019). Leong et al. (2011), for example, present a randomized photo display system to facilitate serendipitous engagements. Strategically using moments of defamiliarization, the system helps people create new connections responsive to their shifting moods or contexts of use. Pensieve (Peesapati et al. 2010; Cosley et al. 2012) creates and sends out email reminders drawing on a personal archive of pictures, recordings, and social media content to invite people to revisit and reflect on them. Odom et al. (2014) introduce a system called Photobox that prints out every month a small number of photographs randomly pulled from a user's online photo collections. Slowing the pace of interaction, the study finds, builds anticipation for rediscovery and reflection, facilitating sustained engagement. While prior research has explored ways of enriching possibilities for revisiting photos by making photos more visible or encounterable, few studies have examined how PVI reminisce with or generally experience their personal photos, or how such photo activities could be supported (cf. Yoo et al. 2021).

5.2.3 The use of mental images by PVI to engage with visual information

While how PVI experience photos remains under-explored in HCI and interaction design, studies have attended to how PVI use mental images to engage with visual information that they cannot directly perceive or access. Kosslyn et al. (2006, p.4) characterize a mental image as something that "give[s] rise to the subjective experience of perception" when "the stimulus is not actually being perceived." The ability to "see" something then is understood as constructing a mental image through which people can "quasi-perce[ive]" and experience something in the absence of actual stimuli (Cattaneo and Vecchi 2011, p.4, p.55).

Prior studies on using mental images have, for example, explored how tactile display techniques could help PVI build mental images of spatial content (e.g., how objects are arranged or are related to each other) to perform tasks such as drawing or map reading (Kurze 1996; Swaminathan et al. 2016). Schaadhardt et al. (2021) studied how blind users experience and work with digital artboards with colored objects through descriptions provided by screen readers. Jung et al. (2021a) found that the usefulness of alt-text descriptions depended on their ability to help blind and low-vision participants "visualize" the information delivered. While prior studies confirm and seek to support PVI's use of mental images, the mental images are typically assumed. We do not understand the relationship between image descriptions and mental images (e.g., how information in descriptions mediate processes of constructing mental images or how the need for constructing mental images affects the information PVI want in image descriptions).

In this paper, we attend to how the mental images that PVI use and construct out of their personal photos shape their expectations for, and experiences of, image descriptions across different photo activities. How could image descriptions support processes of constructing mental images of photos? Would the expectations for image descriptions vary across different photo activities? How could image descriptions be delivered? Developing insight into these questions may help in understanding and designing tools for image descriptions that reflect and enrich how PVI experience their personal photos.

5.3 Data and methods

5.3.1 Statement of positionality

The first, second, and fourth authors of this paper do not experience visual impairments. The third author has congenital low vision and progressive vision loss. The first and the third authors met in 2016 as, respectively, the producer and the main subject of a documentary production on photography and different conceptions of seeing. The first author has since been participating as an observer in online and offline activities for PVI who are interested in photography. Activities ranged from online photo description sessions (where participants posted photos in a chatroom that they wanted others to describe) to guided photo shoots. The second author occasionally participated in those activities also as an observer. The third author has been active in promoting accessible visual media for PVI, participating in photo exhibitions and film festivals as an artist and jury member. While none of the authors are blind, our experiences or observations of living with PVI enabled us to provide rich accounts of how the PVI who we observed engaged with visual information.

5.3.2 Data collection

This paper mainly draws on semi-structured interviews with 30 PVI conducted by the first author. The third author shared a recruitment ad on one of Korea's biggest online platforms for PVI. Our research project was introduced as an interview study of photo use by PVI on their smartphones, including guided tours of photos with permission. The only requirement to participate in the study was having a collection of photos stored on their devices, regardless of its size. Interviews were originally designed to be in person. We decided to conduct interviews remotely via phone or video based on concerns over the COVID-19 pandemic and privacy (e.g., inadvertently sharing photos or other content on their devices that they did not feel comfortable sharing). While conducting interviews remotely did not allow for direct observation of participants' photo use, this format nonetheless enabled participants to describe in rich detail how they themselves experienced and wanted to engage with photos.

The interviews comprised four parts. The first part asked participants to briefly explain their visual impairments and associated conditions. The second part involved a general conversation about their photo activities, device use, and photo collections on their smartphones and other personal devices where applicable. The third part asked participants to describe in detail how they engaged with photos in the following types of photo activities one by one: photo taking, viewing, searching, browsing, sharing, organizing (including downloading shared photos or deleting photos), and editing. For each photo activity, the first author asked participants to walk her through their process of engaging with photos, drawing on specific examples. Where possible and with permission, participants shared some of these photos (or links to their public photo collections) during or after the interview through text messages. The first author asked follow-up questions about the shared photos where necessary, producing additional memos. She also asked participants about any challenges experienced during each photo activity, how they tried to address these challenges, and how they would have liked the activity to have unfolded. The last part of the interview asked participants to reflect on the general meaning of photos, and imagine how they would be or would want to be using photos and image descriptions in the future. We offered online gift cards (30,000 Korean won or roughly 27 US dollars) for participation.

Out of 30 participants, 17 were blind (six congenitally, 11 acquired) and 13 low-vision (six congenitally, seven acquired). 14 identified themselves as female and 16 male. The age ranged between 22 and 64, with the median age of 36 (see **Table 3** for a summary of our participants). Participants could choose between a phone or video interview. Most participants stated that they were not familiar with, or did not feel comfortable, using video chat apps for an interview. 27 interviews were conducted by phone, and three on Zoom. Interviews lasted from 38 minutes to 95 minutes. All interviews were conducted in Korean and audio-recorded.

The first author had also conducted semi-structured, in-person interviews with eight PVI who identified themselves as interested in photography in 2018, focusing on their meaning and understanding of photos. Data from these exploratory interviews was not used in this paper, but reflections on the data informed our analysis of the 30 interviews.

Table 3. Summary of participants and their photo use

Participant	Age	Gender	Visual condition[a]	Onset	Access technology	No. of photos on smartphone[b]
P1	33	F	CTB	Birth	iOS V.O. / Sullivan+	2,000+
P2	60	M	ATB (prev. CLV)	12	iOS V.O.	<500
P3	35	M	ATB (prev. CLV	17	iOS V.O. / FaceTime / Sullivan+	<200
P4	41	M	CTB	Birth	iOS V.O. / Android TalkBack / Sullivan+ / TapTapSee	<100
P5	39	M	CTB	Birth	iOS V.O.	500+
P6	33	F	CTB with some light perception	Birth	iOS V.O. / Sullivan+	1,000+
P7	29	M	CTB with some light perception	Birth	iOS V.O.	<100
P8	45	M	ATB (prev. CLV	15	iOS V.O. / Sullivan+	10-20
P9	61	M	CLV (progressive)	Birth	Android magnifier	1,000+
P10	23	F	CLV	Birth	Android magnifier	2,000+
P11	28	F	ATB (L) / ALV (R)	8 / 10	iOS magnifier / TapTapSee	20
P12	50	M	ATB (L) / ALV (no central vision) (R)	34	Android magnifier, TalkBack / Sullivan+	<100
P13	30	M	CTB	Birth	Android TalkBack / Sullivan+	10+
P14	48	F	ATB	22	Android TalkBack / Sullivan+	800+
P15	57	M	CLV	Birth	Android magnifier, TalkBack / Sullivan+	<10
P16	22	F	CTB	Birth	iOS V.O. / Sullivan+ / TapTapSee	500+
P17	48	F	ALV (no central vision)	30	iOS V.O., magnifier	500+
P18	38	M	ALV (no central vision)	23	iOS V.O., magnifier	30,000+
P19	26	F	CLV	Birth	Android TalkBack., magnifier	20,000+
P20	50	F	CLV	Birth	Android magnifier / Sullivan+	<100
P21	26	F	CLV	22	Android TalkBack, magnifier	2-3,000
P22	64	M	ALV	35	Android magnifier	<10
P23	32	F	ALV (L) / ATB (R)	18	Android magnifier / Sullivan+	500+
P24	36	M	ALV (no central vision)	21	Android magnifier / TapTapSee	2,000+
P25	34	M	ATB (with some light perception)	12	iOS V.O. (Sullivan+ deleted)	<10
P26	24	F	ATB	13	iOS V.O. / Sullivan+ / TapTapSee	2,000+
P27	45	F	CLV (L) / ATB (R)	30	Android magnifier, TalkBack / Sullivan+	800+
P28	38	M	CTB	Birth	iOS V.O ./ Android TalkBack / Seeing AI / Sullivan+ / TapTapSee / Envision Glasses	600+
P29	25	M	ALV (no central vision)	21	iOS V.O., magnifier / Sullivan+	1,200+
P30	36	F	ATB	17	Android TalkBack	<1,000

[a] CTB = Congenital total blindness; ATB = Acquired total blindness; CLV = Congenital low vision; ALV = Acquired low vision.
[b] Estimated number.

5.3.3 Overview of participants' photo collections, motivations, and experience with machine-generated image descriptions

All participants had a photo collection on their smartphones. The size of their photo collections varied widely from under 10 photos to over 30,000 photos. A few participants (e.g., P2, P25) used their digital cameras and computers as their primary devices for photo use, with more photos stored on those devices than on their smartphones.

Three main reasons for their photo use were: (i) capturing information (to identify information or objects in their environment that they cannot directly or immediately see; usually deleted after use); (ii) social participation (to make themselves visible to other people or participate in conversations by sharing or commenting on photos online and offline); (iii) memory-related (to remember not only special moments but also the experience of having been or done something in those moments; photos served as "traces" (P22), "evidence" (P4), or "a medium for reminiscing" (P10)).

23 participants, except for seven low-vision participants who mostly used magnifying features only, were using screen-reading apps (VoiceOver, TalkBack) on their mobile devices. Many participants, however, were not fully aware of how the features of these screen-reading apps could assist their engagement with photos (e.g., VoiceOver and Image Descriptions features on iPhone enable hearing descriptions of what appears in the camera frame). 17 participants were using, or had used, additional image recognition apps (Sullivan+, TapTapSee, Seeing AI) that offered image descriptions. The use of these apps was heavily geared towards accessing textual information contained in photos. Most participants believed that the accuracy of image descriptions would continue to improve given the relative novelty of these apps. There existed divergent views, however, on the usefulness of descriptions. Views ranged from 'any information would be better than no information' (e.g., P4, P28) to 'photos are visual artifacts that could not be properly translated into words' (e.g., P19, P25). While blind participants tended to use image

descriptions more often than low-vision participants, participants' views on image descriptions seemed to be associated with their individual habits and preferences, rather than with the types of visual impairments.

5.3.4 Data analysis

The interviews were transcribed and, where necessary, translated from Korean into English by the first author, who is a native Korean speaker. After each interview, she reviewed the shared photos along with the transcript, producing notes on the photos and reflective memos on the overall interview experience. The transcripts, notes, and memos were combined, and anonymized and lightly edited to reflect the context and improve readability. The initial objective of the study was to explore how PVI in general engage with and experience their photos on their devices. The first author open-coded the initial set of interviews, and the potential themes were discussed with the other authors (Glaser and Strauss 2009). From the first round of data collection and analysis, image descriptions emerged as a potential theme of the study. In the subsequent interviews, questions were revised to probe the role of image descriptions in participants' photo activities in further detail (Briggs 1986). Given the novelty of screen-reading apps on smartphones that generate image descriptions, and participants' lack of experience in using these apps, the questions focused on understanding participants' expectations for (e.g., Stangl et al. 2020; Stangl et al. 2021), rather than their actual experience of, image descriptions. The focus of the questions was on what information participants wanted to see in image descriptions and how they wanted this information to be provided to help them engage with their photos.

In the iterative process of coding and collection (Bernard and Ryan 2010), connecting photos' objective content (i.e., what objectively appears in a photo) to the subjective meaning for participants (i.e., what they think appears in the photo and what it means for them) through image descriptions emerged as another central theme. These themes of image descriptions, photo content, and subjective meaning were further developed in several rounds of data analysis. In analyzing the relationship between these themes, we attended to participants' accounts of their processes of engaging with photos through image descriptions across different photo activities. We also identified a need for participants to construct or modify the mental images that they used to engage with photos in these activities. The concept of "mental images" emerged as our key analytical lens, which we settled on as a way of bringing together our themes in regard to image descriptions, participants' photo activities, and their experience of photos.

5.4 Results: what PVI want in image descriptions across their photo activities

Participants used image descriptions to construct mental images of the photos. In what follows, we characterize how participants wanted to experience photos through image descriptions across four types of photo activities: (i) viewing photos, (ii) taking photos, (iii) searching for photos, and (iv) reminiscing with photos. We explain how expectations for descriptions were shaped by how their mental images varied across photo activities.

5.4.1 Viewing photos

Participants referred to their engagement with photos as "viewing" in the sense that they could visualize the content using image descriptions. To view photos, participants usually started by generating audio image descriptions (using screen-reading or image recognition apps) or by asking their sighted colleagues, friends, or family members to describe photos for them. As P4 commented: "*A sighted person can just see [a photo] and will immediately know what it's about, right?…But we don't have any visual data, nothing's intuitive*." Participants explained that they viewed photos by constructing mental images of photos, or *"a process of putting together and* *visualizing different elements*" (P8). They reported that descriptions "*need to be helpful for visualizing what's there"* (P13). The mental images could be directly about the content of a photo (e.g., P26's example - my dog wearing a Santa hat) or about any aspects of an experience that the participant associated with the content (e.g., the Christmas that I spent with my dog at home). We next identify two types of visual elements that participants said they would need for constructing mental images effectively.

5.4.1.1 Need for spatial cues

Many participants spoke of the need for understanding how the elements of a photo were *spatially* structured so that they could "*put things together three-dimensionally*" (P25) or "*structurally imagine*" (P28) the content. P8 commented: "*If (descriptions) could give me some clues on spatial relations, like, a bird is sitting on the top of a pavilion or a person is standing before the pavilion, it would be helpful for me to imagine the photo more systematically*."

Several participants pointed out that spatial clues in descriptions should logically walk them through the photos, and avoid being "abstract," "confusing," or "fragmentary" (P14). P29 explained: "*Instead of giving me a random description, like, saying 'a person sitting on a chair and a tree and a sky,' I would want the description to read the image...in one direction such as from top to bottom, or from left to right*." In this sense, photo viewing involved participants' ability to construct a mental image of the photo by spatially and logically arranging elements in their heads.

5.4.1.2 Interactively selecting details of photos

Most participants refrained from asking their sighted colleagues, friends, and family members for detailed descriptions of photos to avoid "burdening" them with the task (e.g., P4, P5, P20, P21). The descriptions that they asked for tended to be general summaries (e.g., what kind of photo it is, what the most salient object in the photo is, what's happening) or checks on specific elements (e.g., who is in it, how I look, how this photo is different from the previous one).

To construct a mental image of a photo, participants wanted descriptions at different levels of detail (e.g., P8, P14, P16, P27, P30). For example, they wanted image descriptions to first offer a summary or identify the genre of the photo (e.g., landscape, portrait), then provide information on salient objects (e.g., a tree, three people, and a car from left to right), and then provide more detailed information on each of these objects in the photo (e.g., the faces of three people and what they are wearing).

Participants emphasized the importance of being able to interactively select parts of the photo to generate detailed descriptions. As P20 explained, once an overall description has been offered: "*I want (descriptions) to read just the parts that I want to see…when I tap on them.*" Several blind (e.g., P7, P20, P28) and low-vision participants (e.g., P9, P12, P29) suggested selective details could be provided based on their queries. P29 commented on a photo of a tree and the sky: "*For example, (descriptions) could be really detailed, like, how much of the sky is covered with clouds? Where is the focus placed? Is the background blurry? I could ask things like that*." P28 wanted a general summary ("*the main story of the photo*") to be offered first and then the rest of the elements that "*do not fit the story*" could be offered as individual descriptions that the viewer could explore in an order that they prefer. Five participants (P8, P9, P12, P14, P28) expressed that they would like an option of being able to alternate between overall and detailed descriptions as they like. P12 imagined a system where an overall description is embedded as the background and detailed descriptions of elements would be overlaid so that he could go back and forth as he moved his fingers around the screen, and interactively develop descriptions.

5.4.1.3 Aesthetics and memory

Several participants, especially participants with acquired visual impairments (e.g., P14, P25, P26), connected aesthetic aspects of image descriptions to their ability to remember the photo later. P26 explained: "*Say, there's a photo of the sky. Through image descriptions, it's just the same sky as in any other photos (of the sky). What I want to know, for example, is what kinds of clouds are there, how bright it is, because the blueness of the sky is different every day…The color tones and brightness, things like that can be (linked to) my particular emotions…and [the photo] becomes meaningful and I can remember it better."* Here, descriptions of the aesthetic details of the content of a photo were thought to help produce a more vivid and personally relevant mental image, which would then possibly help the associated photo become more memorable.

P25 shared that his memory of visual information had been fading over the last two decades: "I really try to visualize photos in detail, especially with things like colors, like, 'this will be in this color,' and 'this color doesn't really match with that color.' I keep thinking about things like that in a photo and I think that helps me remember an image for longer." Here, his case similarly indicated a connection between detailed descriptions of aesthetic elements and the ability to construct mental images of a photo that could be remembered.

Taken together, for image descriptions to be useful in photo viewing, participants expected them to offer clues about how the visual elements in photos were spatially arranged, to provide varying levels of detail that could be selected interactively, and to describe aesthetic elements to make photos more memorable.

5.4.2 Taking photos

Participants took photos themselves, or delegated photo taking to sighted persons (especially in the case of blind participants). Photo taking began by constructing a mental image of the intended photo (e.g., P6's example - an image of a dog sleeping on my lap) based on what they were experiencing (e.g., I'm at a dog cafe and a dog just jumped onto my lap) or descriptions by sighted people (e.g., "the dog just closed its eyes"). To take the photo, participants compared the image of their intended photo with that of the camera frame until the latter was close enough to the former.

5.4.2.1 Taking the photo they intended

Most participants found it difficult to operate their smartphone cameras to produce objectively "good," "acceptable," or "usable" photos (e.g., "*the photos [the sighted] take would be better than one that I take"* (P1)). The more important goal, however, was whether the photo reflected what they had intended (e.g., "*it's not about being able to tell whether a photo I've taken is a good photo, but about whether the photo captured what I wanted to capture"* (P26)). All participants reported frustration over the gap between their intended images and the actual content of their photos, which often deterred taking photos in the first place. As P4 put it: "*The reason that I don't really enjoy taking photos is not simply that I can't see, but that I keep getting this interference with the photo I imagined...Taking photos then becomes no longer meaningful and my effort just gets wasted"* (P4). Reducing the gap between the actual photos and their intentions also mattered for using image descriptions. P18 commented: "*This may sound like a chicken-or-egg problem, but you need to be able to capture what you want in a way you want so that you can get descriptions that you want."*

A few of our participants (P3, P28) were using tools (e.g., apps such as VoiceOver or Seeing AI) that offered image description features for photo taking such as real-time audio feedback (e.g., how many faces are detected and where they are located in the frame) for framing a photo. The real-time audio feedback, however, was often insufficient to be useful in bridging the gap between their intended images and the content of photos. Our participants surfaced two specific challenges for image descriptions in this regard

5.4.2.2 Spatial cues to iteratively construct and adjust the actual content

As when viewing existing photos (5.4.2.1), participants wanted image descriptions to provide clues on spatial relationships between objects in their camera frames when taking photos. Their desired use of spatial clues for photo taking was not only to help them construct a mental image of the objects in the camera frame, but to adjust the frame itself. This extra task of needing to figure out what the camera frame shows led to a desire for more detailed spatial clues (relative to 5.4.2.1). Participants wanted to know not just which objects are in the frame and where they are positioned, but also how much space each object takes up, how much bigger one object is versus another (e.g., how big a tree is versus a person, or how big my eyes look compared to my whole face), and how much of an object is in the frame (e.g., whether I captured the whole tree, or only a few branches). It was important for participants to be able to use descriptions to adjust what was in the camera frame both before and after taking a photo. As P28 explained: "*Sighted people see something first and decide whether they want to take a photo of it…and how they will frame it, but it works almost in reverse order for us [the visually impaired]...We have to take a photo first only to discover what's wrong with it."*

Several low-vision participants spoke of dividing the descriptions of the frame into a few numbered regions to help them adjust the contents in the frame more precisely. P18, for example, suggested the use of quadrants in descriptions: "*As long as my kids don't run around, I usually tell them 'stay right there' and try to roughly adjust the camera. If descriptions can tell me in which quadrant my kids are, that would be great.*" P27 suggested descriptions may be based on nine numbered "grids" to help her configure the image more in detail: "*I might, for example, want more leaves (in the frame), like in the top-left four grids, and put something in (grid) number 8*."

A few congenitally blind participants (e.g., P5, P7, P16) saw descriptions of spatial clues as potentially useful for learning to take their intended photos themselves. P16 shared that she "*practices*" with iPhone VoiceOver descriptions on the position of her face in the front-facing camera frame (e.g., adjusting her arms according to descriptions such as "*one face near top left edge, in the middle*") to get a sense of how far or high she needed to hold her smartphone to produce the selfies she wanted. P5 talked about practicing over the years to better capture his intentions in photos by trying to take photos of the same object (e.g., still objects like a cup) from multiple angles and distances, and asking his sighted friends to describe how these photos were different. He added: "*People say 1cm or 1 degree can change a photo a lot and I want to experiment more in detail...what I'll be able to express so that I can take photos that I want.*"

5.4.2.3 Feedback on how less salient elements appear in the frame

Many participants wanted image descriptions to capture elements that might not be visually salient, but which they viewed as important (e.g., two people looking at each other, people with their eyes open, a shape of clouds, a snake in a particular shape, or leaves reflecting the sunlight). P2 recounted when he traveled to a beach with his guide dog: "*I felt that [my dog] was gazing at the sea, so I wanted to take a photo of him doing that*." He took multiple photos behind his dog only to learn later from his wife that his dog was not looking at the sea in any of these photos.

Aspects that participants wanted to capture in the frame varied widely, both across participants and in the contexts of photo taking. A few participants questioned the feasibility of generating descriptions that could capture what was important, given the inherently subjective nature of photos and the challenges of communicating to description tools what was important in a particular photo ("*nobody would know what I think is important in that photo, machines won't know, will they?*" (P19)). Finally, in terms of interaction, participants wanted confirmation of descriptions generated by the participants themselves (e.g., Are the two people looking at each other? Is the snake curled up in a circle?).

In summary, image descriptions for photo taking were expected to help in adjusting the camera frame to get a photo close to participants' intended image. Participants reported that use of image descriptions was challenging since current tools did not provide descriptions of spatial clues suitable for adjusting the camera frame, or for confirming how elements that were not visually salient but subjectively important for them appeared in the frame.

5.4.3 Searching for photos

Most participants expressed that image descriptions were potentially useful for searching photos on their devices. To search, participants would go to their photo galleries on their devices and navigate through their photos with screen-reader apps that generate descriptions, including image types (e.g., photos, screenshots), date, time, and location data, or the main objects or text in photos. For example, when a participant enters their photo gallery with a screen-reading feature turned on, they would search for a photo by tapping on any part of the screen to hear a description (e.g., "photo, April 2, a cat laying on top of a speaker, Marshall"; "screenshot, April 2, a screenshot of a map with text and numbers"). In some cases, participants would describe their desired photo to a sighted person, who would try to find a photo that matched their description. A few participants (e.g., P2, P4, P18, P26) tagged or renamed particular photos, which they used for searching for these photos later.

5.4.3.1 Actual photos different from what was remembered

Participants' memory of photos that they wanted to search for (e.g., the mental image that they had stored) was often different from the actual content of the photos. They might remember the date or location but not the content of the photo, or they might remember the content in detail without remembering when it was taken. Such cases were equally common across the blind participants (e.g., "As I was not taking a photo of something that I'd visually seen, the [actual] photo [I searched] could be completely different from [the photo] I pictured in my head" (P28)), as well as the low-vision participants ("I thought the flowers were in full bloom but it turned out I took a photo of empty branches" (P29)). Differences in the photos that participants wanted and the actual content of the photos that they searched for often caused frustration, whether searching themselves or with the help of a sighted person. P30 commented on challenges when asking her sighted helper to find photos for her: "I usually describe [to the helper] a photo and then she'd find something similar and describe it back to me…Her description would be different from what I remembered but I know my memory could be wrong…then there's no way that I can confirm that is the photo that I was looking for." In such cases, descriptions of the objects in a photo, however detailed, may not be helpful for participants to search for photos.

5.4.3.2 Hard to differentiate similar photos

Participants frequently searched for a photo based on the date and location that they remembered it was taken. Yet searching for the right photo often required additional clues. P12 explained: "*When I turn VoiceOver on and go to my photo gallery, it tells us the date and time (photos were taken). But if I had taken multiple photos of a similar thing, say, I took a lot of photos with my daughter around a park, I cannot find a specific photo without detailed descriptions of the background - Where exactly in the park was this taken?*" P1 similarly recounted difficulty in retrieving a photo with generic descriptions even when she knew exactly when and where the photo had been taken: "*I wanted to show (my friend) a photo of a snake curled up in a circle…The photos that I took that day with the snake will be all slightly different…but VoiceOver would just read them all 'a female, a snake'.*" As in 5.4.2.2, searching for a photo could require detailed descriptions on aspects of photos that are not visually salient (i.e., items in the background or the shape or motion of the snake) to enable differentiating among photos of similar content.

Many participants suggested that details might be searched as "keywords," which they could use to filter photos in a manner almost akin to web search or faceted search (e.g., Tunkelang 2009). For example, P29 explained how he might search for a photo based on keywords: "*Say, I went to the beach with my friends. We'd be wearing (in the photo) short sleeves if it were summer, overcoats or padded jackets if it were winter. Then I could just enter 'beach,' 'overcoats,' and other related keywords to find the photo*." P18 similarly commented on his strategy to find older, everyday photos of his children: "*I first guess how old [my kids] must have been, and what they were wearing, and their hairstyle…to successively filter photos based on these little pieces that I remember*." Descriptions offered as a list of words were not perceived by participants as providing adequate spatial clues for viewing or taking photos (5.4.2.1 & 5.4.2.2). In photo search, however, keywords were seen as potentially effective. The reason was that, whereas in photo viewing and taking, words referred to the objectively identifiable visual objects in the content, in photo search, keywords referred to clues in a subjective mental image that were used to evaluate other photos against the mental image.

5.4.3.3 Use contexts can matter as much as the photo content itself

In the previous two sections, participants wanted to search using content-related clues about a photo such as the date, location, and other spatial clues. Participants also often relied on the contexts surrounding using a photo (e.g., who sent me the photo, where it is stored, relevant conversations). Several blind and low-vision participants (e.g., P1, P14, P16, P18, P19, P20, P24), for example, searched for photos on their chat apps instead of browsing through photo galleries on their devices. P19, for example, explained how she used her chat app to search for a frequently used profile photo: "*If I had sent that photo to someone, I track that activity in that chat room to find the photo. Then I download the photo into my photo gallery again, and send myself a message (on the chat app) containing the downloaded time and date and a short description (e.g., May 1, 3pm, "profile photo")…so that the next time, I can just search for the text "profile photo" (on the app) and get the time and date, and then go to the photo gallery to find the photo with that information*." Other participants similarly used their chat apps to create and collect use data about photos that could be more easily searched for and retrieved as texts. Where descriptions on the content of photos alone were not adequate to locate a photo, participants created and leveraged data about its use contexts as additional clues.

Taken together, participants encountered several challenging situations in searching their photos. Search was challenging in that the actual photo was often not how they remembered it, descriptions of the content were not detailed enough to differentiate similar photos, and clues about the use context were needed beyond the content itself to remember and retrieve a photo.

5.4.4 Reminiscing with photos

All participants understood photos as important for reminiscing: "Through photos, I just want to record something as I intended and then reflect on them on my own, without others' help" (P30); "I take photos…then they can remain in my memory" (P5); "As long as I photograph (something), I can remember it later" (P16). As P25 put it, participants wanted to use photos as "memories in an album" that could be "revisited and revived," rather than as "just files stored in a folder." To reminisce with a photo, participants would listen to a description of a photo to trigger a memory of a moment or experience: "Sometimes my kids would play with my phone and find their baby photos. When they describe the photos for me, I'm like, 'Ah, you were so cute then! That is what it was like then, yes!'...[the photos] are like our shared memories" (P14). P2 liked to browse through the photos he had taken and use descriptions to reminisce. He did so both with the help of his sighted wife, who would describe the content of photos for him, and by himself using the "file names" of photos (usually in the format of "location/date/a short summary (e.g., my dog looking at the sea)") that he created when he reviewed his photos. The descriptions helped him to visualize the moments of photo taking through which he could "bring back memories."

Some participants often searched for photos to reminisce about associated experiences. The descriptions that they encountered as they searched and viewed the photos helped them remember these experiences. P23 commented: "*I've told you that I traveled around Eastern Europe in 2019, right? But, in fact, I don't really have a strong memory of that trip. So I often search for those photos…the (Sullivan+) app kind of tells me what's there and even with that I could recall what I did then…Rather than looking at the content itself, I revive my memories through the presence of a photo*." Despite their desire to reminisce with photos, participants faced several challenges, which we explain next.

5.4.4.1 Image descriptions not connected to memory

Image descriptions often failed to elicit or trigger a particular meaning or memory. P26 illustrated such a case with a photo of an alleyway: "*When a sighted person sees a photo of an alleyway, they would go right away 'ah, I used to live near there.' But as a visually impaired person, you wouldn't get a description much more than 'a selfie in a rundown alleyway,' would you? Then, 'so what? What's that to do with me?'...I want my photos to be directly connected to* *my memory."* To help her connect the descriptions to her memory, she suggested that her personal data stored on her device(s) could be included in the descriptions. As an example, if she takes a photo at a location, the description could not only provide information on the content but add "you used to live here for X years." This example indicates how reminiscing with photos by PVI may require including descriptions of how the photos' content is relevant to the person.

Interestingly, participants reminisced also by connecting a photo to new mental images. New connections often happened in collaborative photo viewing activities. When P5 took a photo of the piano in a war memorial, different people provided him with different descriptions: "*Someone said: 'there's a piano and next to it is something' and another person said 'the background is white but the piano is in the shade, and it seems like it will play sad music.' [These different descriptions] can be confusing, but I think they are in a sense complementary*." Here, different mental images of the same photo were constructed according to different descriptions. P1 also said that she liked looking at photos with people who had not been with her when they were taken, as they helped her re-experience her photos differently: "*The way I feel about a photo came from what I experienced at that place, at that moment…people describe a photo, then I realize that's another way of looking at it. I can not only revisit my memory but I also feel that my memory has become richer*."

5.4.4.2 Conjuring up what it was like

Participants wanted to be able to conjure up what it was like to be there at that moment of photo taking. Several participants pointed out that constructing a mental image that brings back what it was like depended on the ability to elicit the atmosphere of a photo (P14, P23, P25, P26, P28), or how they would have been feeling (P16, P21, P26, P27, P28, P30).

A few participants (e.g., P19, P21, P27, P28) commented that descriptions to support reminiscing with photos would need to go beyond factual content. P28 explained: "*AI-generated descriptions are at least for now mostly factual (that tell us) what's there…Such descriptions might help me see 'ah, ok, it's a photo of something,' but would they be able to help me recall the meanings that photo contains…or realize 'ah, yes, I've seen that before'?"* Other participants cautioned against including potentially subjective adjectives (nice, pretty, etc.) in descriptions as they wanted to be able to determine the atmosphere of a photo on their own. In this context, P14 explained that she tried to ask for descriptions from someone who shared her aesthetic preferences so that she would be able to evaluate a photo in a way that reflected her ways of looking at things. She also suggested that image description tools could collect data on the types of visual content that she liked (e.g., 'liked' photos, colors or objects captured frequently in her photos) and incorporate such data into their descriptions.

P26 commented that a useful image description of a photo's atmosphere would require a combination of a wide range of details: "(Descriptions of) brightness, facial expressions, the setting - is it a restaurant or a hiking trail? What are people wearing? If I put them all together, I would be able to visualize or imagine the atmosphere, and relive in that moment." Several participants (e.g., P2, P3, P16, P24, P26, P28, P30) compared their desired descriptions with the sound in their video recordings, which was more accessible or "easily recognizable" for them, helping them construct mental images of an atmosphere more "intuitively and instantly" or "vividly." P28 explained: "Descriptions cannot tell us everything on the screen, but sounds could capture an idea of 'it is this kind of thing,' 'it would feel like this.'" He suggested, where possible, image descriptions may accompany sounds that matched a photo's content (e.g., adding ambient sounds of a cafe to a photo taken at a cafe, or sounds of rustling leaves to a photo of a street in autumn). Participants indicated that conjuring up "what it was like" was a rich multisensory experience that image descriptions would be able to capture only partially.

5.5 Discussion

5.5.1 Image descriptions to support engagement with personal photos by PVI

In **Table 4** below, we summarize our findings. For each photo activity, we describe the key process involved and the content of image descriptions that our participants expected, along with design opportunities and relevant prior work.

Table 4. Summary of findings

Photo activity	Key process	Expected content of image descriptions	Design opportunities
Viewing photos	Constructing a mental image of the content of a photo.	• Spatial clues on visual elements provided in a consistent order • Varying levels of detail that could be selected interactively • Aesthetic elements to make photos more memorable	• Design descriptions that can logically walk the user through the photo (left -> right, top -> bottom) (e.g., Stangl et al. 2018). • Design descriptions at different levels of abstraction and which allow for interactive exploration (e.g., Morris et al. 2018). • Include more aesthetic descriptors.
Taking photos	Constructing and adjusting a mental image of what appeared in the camera frame, then comparing the image to that of an intended photo.	• Spatial clues that help the user iteratively construct mental images of the content in the camera frame and adjust the camera frame • Clues that help the user confirm queries about non-salient elements in the frame	• Offer detailed real-time descriptions of the content in the camera frame (e.g., using both predefined and user-defined regions). • Generate descriptions based on the user's queries (e.g., Morris et al. 2018).
Searching photos	Finding a photo whose description matches the mental image that the user has in mind.	• Additional clues about the photo in case the actual photo differs from the mental image • Details that help the user differentiate similar photos • Data on how a photo was used on the user's device	• Include data on both the content of a photo and contexts around the photo drawing on activities on the device and other types of metadata. • Generate descriptions based on the user's queries (e.g., Morris et al. 2018).
Reminiscing with photos	Constructing a mental image of a photo to trigger, relive, or enrich an associated memory.	• Personal data that helps the user connect a mental image to memory • Aesthetic details that help the user "conjure up what it was like"	• Include personal data that is relevant to a photo. • Offer a range of multisensory data that the user could request on demand (e.g., Morris et al. 2018; Lim et al. 2019; Rector et al. 2017).

Extant work in HCI and accessibility has mostly studied the use of image descriptions by PVI when engaging with publicly available online images for functional purposes (e.g., reading textual information captured in a photo) (Kindberg et al. 2005; Bennett et al. 2018b). Extant work explores, for example, what content to include in descriptions to help the photo viewer identify or recognize objects in the photos. The motivation of our study was to explore what PVI want in image descriptions to support their broader engagement with personal photos. In our study, participants wanted to engage with their personal photos not only for functional use but also to talk about experiences with others, and to reflect on or reminisce about their past moments (Zhao et al. 2017; Harada et al. 2013).

By exploring the use of photos by PVI on their personal devices, our findings highlight how image descriptions to support personal photo use are in several important ways distinct from image descriptions for publicly available images or for functional purposes. To search for their personal photos, for example, we found that even accurate and detailed descriptions of visually salient objects could be inadequate since participants' mental images of the photos often differed from their actual content. Similarly, image descriptions needed to be connected to participants' memory (e.g., when and why they took the photo) to be useful for reminiscing. In these cases, descriptions of the content of photos, however detailed, would not be able to fully support their intended photo activities.

Our findings also indicate that image descriptions need to account for the contexts of photo use. A few scholars have begun to explore how the content of image descriptions that PVI expect is context-dependent. Stangl et al. (2020), for example, examined how the descriptions expected by PVI varied according to the source of the online image, such as whether the image was from a news site, social media platform, eCommerce site, or online publication. Stangl et al. (2021) further investigate how the expected content of descriptions also depended on “scenarios” that shaped the purpose of using image descriptions (e.g., learning from a news site, planning a trip on a travel website, or sharing photos with friends).

Extending this prior work on publicly available images, our study explored how PVI want to use image descriptions in activities of viewing, taking, searching, and reminiscing with their own photos on their personal devices. The descriptions that participants wanted differed across their photo activities. For example, while participants wanted to get spatial clues of visual elements when viewing and taking photos, they expressed a need for more detailed spatial descriptions when taking photos as they wanted. In taking photos, they needed not just to construct a mental image of the content in the camera frame, but to make the image close enough to that of their intended photo. As another example, descriptions offered as a list of words were not perceived by participants as providing adequate spatial clues for viewing or taking photos. When searching for photos, however, lists of “keywords” were seen as potentially effective. Also, expectations for aesthetic details and personalized descriptions were more common in photo activities for reminiscing. By surfacing how the image descriptions that participants wanted varied by their photo activities, our study adds to existing insights on how to structure and deliver image descriptions (e.g., Morris et al. 2018; Stangl et al. 2018).

Drawing on the opportunities presented in the right-hand column of **Table 4**, we propose three considerations when designing image descriptions to support personal photo use of PVI that goes beyond identification of objective content: (i) support for PVI to construct mental images of photos; (ii) accounting for how the intentions and goals of PVI vary across photo activities; and (iii) exploring the relation between image descriptions and memory.

5.5.1.1 Support for PVI to construct mental images of photos

Participants did not just experience image descriptions as a list of objects or other visual elements contained in a photo, but as clues to be arranged or put together in a particular way to construct a mental image through which they could experience the photo (e.g., Jung et al. 2021a). Clues could concern the content of a photo or any aspects of an experience associated with the photo. By highlighting how personal photo use of PVI depends on their ability to construct mental images of their photos, and by unpacking the different processes of constructing mental images, our study points to design opportunities for going beyond identification of the types of data to include in image descriptions to supporting the constructive processes by which such data is pieced together.

5.5.1.2 Accounting for how the intentions and goals of PVI vary across photo activities

The process of constructing mental images of photos varied across photo activities, for which participants had different intentions. Identifying such differences contributes concrete design considerations for generating image descriptions by photo activity (e.g., structuring descriptions to walk the user through the content when viewing photos; allowing for query-based descriptions when searching photos; and providing descriptions about both the content and the context of photos (personal data, user activity data, and other metadata) when searching or reminiscing with photos). These considerations also indicate a need to support PVI to define and account for their own intentions and goals. For example, while research on photo support for PVI tends to suggest designer-defined guidelines for how to take “good” photos (e.g., content framed in a particular way), participants considered it more important to take the photos that they themselves intended.

5.5.1.3 Exploring the relationship between image descriptions and memory

PVI associated photos with memory. It follows that image descriptions need to help PVI remember photos. Extant literature on image descriptions and photo support for PVI, however, has given less attention to what facilitates or limits the use of descriptions for memory. In our study, participants expressed that aesthetic details helped them construct more vivid, emotional, and thus more memorable mental images of photos. As mental images could concern both the content of and the context surrounding a photo, our findings also indicate that descriptions may need to go beyond faithfully describing the content of the photo. Descriptions could bring in a rich range of multisensory data on the user's personal life that could be relevant or connected to the time at, and location in which the photo was taken (e.g., whether the user lived near where this photo was taken and for how many years, or what kind of music the user listened to in the month that this photo was taken). Given that such support risks generating irrelevant connections between data (e.g., Odom et al. 2014) or connections that the user could feel uncomfortable with (e.g., memory of an ex-boyfriend), the user would need to be given flexible control over what kinds of data could be used for generating image descriptions.

5.5.2 Theoretical implications for understanding the role of image descriptions for PVI

The findings and design opportunities that we summarized in 5.5.1 also hold broader theoretical implications for HCI research on the role of image descriptions in experiencing photos. Scholars of the phenomenology of images emphasize that what matters in photos is less the content itself than people's subjective experience of engaging with the content (e.g., Barthes 1981; Husserl 2006; Pettersson 2011). Rather than just seeing "something is there," we experience "I have been there," or "what it would be like to be there." Such experience of photos is multisensory (Pink 2011) (e.g., looking at a photo of a beach helps me experience what it smelled, sounded, or otherwise felt like to when was I there). Likewise, our findings indicate that PVI understand personal photos not simply as visual content to be objectively and faithfully translated into words, but as a multifaceted artifact that they can subjectively and selectively *experience* (e.g., exploring elements of photos at varying levels of detail, connecting some aspect of the content to personal memory, conjuring up what it would be like to be in the moment depicted in the content). Participants' photo activities for reminiscing also highlighted the potentially multisensory experience of photos (e.g., Webster and McCall 1999).

In contrast, much research in computer vision aims to understand and reproduce humans' ability to identify and describe the content of images (Gurari et al. 2020; Hossain et al. 2019). Many image captioning systems, for example, have been designed to narrow the gap between descriptions of machines and sighted people detecting and identifying visual elements (MacLeod et al. 2017). To support asking questions about the content of an image, it has been similarly assumed to aim for producing a single answer (Bhattacharya et al. 2019). While it is important for image descriptions to allow PVI to access the content of a photo that they themselves do not have direct or immediate access to, we have shown that photo activities of PVI required descriptions to help them go beyond identifying the content of their photos.

The broader theoretical implications are that the role of image descriptions for the photo use of PVI might be understood as supporting the multisensory experience of what a description of visual content triggers, rather than descriptions of the objective content to substitute for missing or lacking visual input. Design considerations for image descriptions for multisensory experience could extend and integrate research in HCI and interaction design on multisensory systems for both PVI (e.g., Yoo et al. 2021; Yoo et al. 2020; Rector et al. 2017; Morrison et al. 2017; Adams et al. 2016) and the sighted (e.g., Fennell and Frohlich 2005; Frohlich 2004).

5.5.3 Implications for designing image descriptions for the sighted

Our findings about PVI may also be used to inform the design of image descriptions for the sighted. The photo activities of PVI share some characteristics with those of the sighted. Sighted photo users similarly often do not remember what photos that they have on their devices, and search for photos using data about photos' contexts (e.g., the date, location, or app on which the user has shared the photo before). Studies have also found the need for personalized descriptions to support sighted users to reminisce with their photos (e.g., Peesapati et al. 2010; Cosley et al. 2012).

With the dramatic increase in the number and types of photos that people take, share, and manage on their personal devices, supporting meaningful and continued engagement with personal photo collections has become an important design challenge (Petrelli and Whittaker 2010; Odom et al. 2014; Van House 2016; McGookin 2019). We believe the elements of image descriptions that we surfaced in our study could also be drawn on as design support for this challenge. For example, using a range of multisensory data relevant to a photo (e.g., providing a short description of a photo with the music most frequently played when it was taken) may also help sighted users to reminisce by experiencing the same photo differently or more richly over time. Such considerations also relate to emerging research on generating captions that are more personalized and engaging, rather than repeating what is visible (e.g., Shuster et al. 2019; Park et al. 2017). This is just one example of how studying personal photo use of PVI may allow us to design descriptions that are not just about what is literally in the photo, but enable us to experience the world with what Cattaneo and Vecchi (2011) referred to as "an *other vision*" (p.206, emphasis original).

5.5.4 Limitations and future research

Our study seeks to open up opportunities to design personalized image descriptions to help PVI meaningfully experience and engage with personal photos on their devices. We acknowledge that not all the recommendations we make in **Table 4** (e.g., providing rich aesthetic detail or multisensory data on demand) may be currently feasible. We believe, however, there are several possibilities for exploring the feasibility and design requirements of image descriptions that we suggest. For example, we may conduct a Wizard of Oz study that lets participants explore different parts of a given photo across different photo activities to identify sequences or patterns in participants' use of descriptions. Further research could also be conducted on which data or metadata that participants want (or do not want) to be incorporated in image descriptions of their personal photos, and whether such use of personal data raises privacy concerns.

Our study built on phone or video interviews with 30 participants with vision impairments. While the interview data combined with shared photos revealed rich details of how participants engaged with their photos across different photo activities and what they expected from image descriptions, we did not directly observe their engagement with photos on their devices. Future work may examine participants' photo activities in person (e.g., using contextual inquiry or participatory design methods) to develop additional design considerations that might not have been possible to identify through phone or video interviews. All our participants were Korean and, at the time of our study, there existed very few image recognition apps that supported Korean language users. We need more work on how participants' exposure to such tools may change their expectations for image descriptions.

We focused mostly on how participants experienced photos individually as they related to constructing mental images and connecting them to personal memory. Many of our participants reported that they also engaged with their photos with the help of other sighted people. How such interaction happened, however, was not explored in this study. Future work may examine the interdependence or partnership between PVI and the sighted and how that mediates the role of image descriptions (e.g., Bennett et al. 2018a; Vincenzi et al. 2021).

Lastly, in our attempt to surface common patterns across participants to explore design opportunities, we did not analyze in detail differences in blind and low-vision participants' experiences with photos (Zhao et al. 2017). A systematic comparison between blind and low-vision participants may help provide image descriptions that are tailored to the needs of people with different vision conditions.

5.6 Conclusion

This chapter drew on interviews with 30 PVI to surface how they engaged with their personal photos and their expectations for image descriptions. The study found that the mental images that participants constructed to engage with photos, and the content of image descriptions that they expected, varied according to whether they were viewing, taking, searching, or reminiscing with photos. The findings point to a need for image descriptions generated to support PVIs' personal photo use that goes beyond functional purposes. I suggested that image descriptions for PVI may provide support for constructing mental images of photos, accounting for different intentions and goals of PVI across photo activities, and exploring the relationship between image descriptions and memory.

The study showed that participants did not just experience image descriptions as an objective list of objects or other visual elements contained in a photo, but as data to be arranged or put together in a particular way to construct a mental image through which they could engage with the photo. Data could concern the content of a photo or any aspects of an experience associated with the photos. By highlighting how personal photo use of PVI depends on their ability to construct mental images of their photos, and by unpacking the different elements of a photo for constructing mental images, the study points to design opportunities for going beyond the identification of the types of data to include in image descriptions to supporting the constructive processes by which such data are pieced together.

This study also relates to research on the phenomenology of images. This research emphasizes that what matters in photos is less the content itself than people's subjective experience of engaging with the content. Rather than just seeing "something is there," we experience "I have been there," or "what it would be like to be there." My findings likewise indicate that PVI understand personal photos not simply as visual content to be objectively and faithfully translated into words, but as a multifaceted artifact that they can subjectively and selectively experience (e.g., exploring elements of a photo at varying levels of detail, connecting some aspect of the content to personal memory, conjuring up what it would be like to be in the moment depicted in the content). This study thus highlights the potential usefulness of a phenomenological view of data when designing image descriptions to support PVI's experience of photos.

6 How Artifacts Under Repair Become Contingently Stabilized

In this chapter[4], I present the final study on data users' data practices. The focus in this study is on the inherent multiplicity and uncertainty of the artifacts with which data users engage, and the processes through which such multiple and uncertain artifacts are turned into actionable versions.

Drawing on a series of vignettes from ethnographic fieldwork at an analog electronics repair community in Seoul, South Korea, the chapter focuses on processes through which repair workers and owners negotiated the "repairedness" of a radio, audio amplifier, and electric pipe organ. By "repairedness" I refer to the contingently stable, working version of an artifact under repair that is negotiated out of multiple possible versions to bring about the temporary conclusion of repair work. The study unpacks how the set of properties of artifacts under repair were identified, prioritized, and evaluated as criteria for these artifacts to achieve repairedness, or a sufficient state of repair. These properties were also scoped, re-prioritized, and temporarily substituted for throughout the repair process, shaping and re-shaping how a given artifact under repair was supposed to look and function for it to be considered working again.

While the study does not explicitly mention "data," repair workers' engagement with properties of artifacts can be usefully understood as how they selected and put together data about those properties in their given situation (e.g., reading and analyzing numbers on test meters as data

[4]This chapter has been published as Jung et al. (2021b). I proposed and conducted the ethnographic fieldwork. I analyzed the findings and wrote the manuscript with feedback from the co-authors (Tom Steinberger, John L. King, and Mark S. Ackerman).

on particular electrical properties of a machine (voltage, current, etc.), or evaluating the sufficiency of the volume level of a radio or the boot time of an electric pipe organ). Studying processes of negotiating the repairedness of an artifact highlights the contingency in the properties of the artifact itself, and the need for repair workers to work with data about the properties of an artifact under repair that are experienced as relevant to a given repair situation to settle on one actionable, working version of the artifact contingent on the situation.

6.1 Introduction

Studies of repair have been an important area of research in CSCW as well as STS and HCI (Jackson et al. 2012; Jackson 2014; Rosner and Ames 2014; Mikalsen et al. 2018; Graham and Thrift 2007; Denis et al. 2016; Houston et al. 2016; Maestri and Wakkary 2011). Drawing contrast with a view of technologies as stable designs, this research emphasizes contingent and ongoing processes of performing repair work, and of the valuation of artifacts under repair (Orr 1996; Suchman 2007; Büscher et al. 2009; Houston et al. 2016). As Jackson (2015, p.233) notes, these contingent, ongoing processes of repair have been a key source of insight in CSCW into how sustainable engagement with artifacts depends on going beyond the "thin functionalism" implied by their *a priori* design. Examples of these contingent, ongoing processes include diagnosing breakdowns in artifacts (Orr 1996; Büscher et al. 2009), negotiating the scope and meaning of a repair job by mobilizing networks of stakeholders, practices, knowledges, and infrastructures (Rosner and Ames 2014; Dye et al. 2019; Mikalsen et al. 2018; Wolf et al. 2019), and modifying or repurposing artifacts under repair to fit shifting contexts and environments (de Laet and Mol 2000; Jackson and Kang 2014; Wyche et al. 2015).

While studies of repair have surfaced processes of making problematic artifacts work again as an "ongoing, frequently artful, and often fraught accomplishment" (Jakson 2017, p.174; also

Houston et al. 2016), few of these studies make explicit how and when these processes are concluded. That is, while we have diverse insights into the processes of "getting a repair job done" (Orr 1996), we have less understanding of the processes of negotiating and determining what constitutes a done job. When does an artifact under repair become at least contingently stabilized as working again? The relevance of this question to broader issues of both repair and sustainable engagement in CSCW is suggested by several STS studies (e.g., Law 2002; Mol 1999; Mol 2002, p.43) on how artifacts are not just historically contingent but have a "complex present" that depends on negotiating and enacting a single version out of their multiple possible versions. Emphasis on the multiplicity of artifacts points to what one might call the *thick* functioning of the artifact under repair — its many micro-level properties that affect its performance, and which variously become salient or fall into the background according to a given repair situation. A multiplicity lens suggests that framing the process of stabilizing an artifact under repair into a working version can bring attention to micro-level properties of artifacts that have been less examined in CSCW research on repair.

Consider a used smartphone. Its screen is partially cracked and darker than when it was new. It frequently loses WiFi connections but stays connected to a local network. Apps on the device freeze sometimes. When fully charged, it would last less than a few hours, even when not in use. Does this device work? Is it in need of repair, or is it broken? Now, imagine you take it to a smartphone repair shop. The repair technician fixes a power problem but also offers to fix other problems. Is the device now working and repaired, or is it working but still in need of repair? The process of concluding the work of repairing this smartphone points to the non-binary nature of repair work and its outcomes (e.g., Houston et al. 2016; Cohn 2016). Rather than becoming fully repaired or objectively working, the properties of the artifact under repair may take on multiple possible versions of workingness that the owner and repair worker need to negotiate. A single working version of the smartphone is contingently arrived at by negotiating these properties.

In this paper we refer to how repair workers and owners negotiate a working version out of the multiplicity of an artifact as the process of negotiating a state of *repairedness*. Our paper seeks to develop a concept of repairedness in order to make two main contributions to CSCW. First, we seek to frame the negotiation processes underlying our concept of repairedness as a lens on the less-studied conclusion of repair work. Second, we seek to link the theoretical implications of repairedness regarding the multiplicity of artifacts to broader conversations in CSCW as well as STS and HCI on the role of artifacts in collaborative work settings (e.g., Lee 2007; Schmidt and Wagner 2004; Cabitza et al. 2013; Muller et al. 2016; Ackerman et al. 2013) and sustainable engagement (Meurer et al. 2018; DiSalvo et al. 2010; Jackson 2014).

To develop these contributions, this paper describes a series of vignettes from an ethnographic study of an analog electronics repair community in Seoul, South Korea. The professional repair workers dealt with a wide range of idiosyncratic, mostly analog machines that found their way into the community after being rejected as unfixable elsewhere. Our site thus offered rich insight into how people engage with multiple versions of artifacts under repair. Our analog electronics repair setting also adds to the diverse research settings in CSCW repair studies that span digital electronics (especially mobile phones) repair markets in the Global South (e.g., Jackson et al. 2012; Ahmed et al. 2015; Wyche et al. 2015; Houston and Jackson 2016), amateur repair communities (e.g., Rosner and Ames 2014; Houston et al. 2016), repair projects at home or for art (e.g., Maestri and Wakkary 2011; Wolf et al. 2019; Jackson and Kang 2014) as well as infrastructure projects (e.g., Cohn 2016; Mikalsen and Monteiro 2018; Dye et al. 2019).

In our vignettes, the repair workers and owners negotiated versions of the artifacts under repair by strategically identifying and prioritizing particular properties of the artifacts as well as determining the sufficiency of these properties in a given repair situation. These properties were also scoped, re-prioritized, and temporarily substituted for during the negotiation process, shifting what constituted a working version of the artifact under repair. Repairedness, rather than being an *a priori* state of the artifact, was a contingent state that emerged from these negotiation processes.

In what follows, we review existing studies on repair work in CSCW and the multiplicity of artifacts in STS, and lay out a concept of repairedness by bridging these studies. We then provide an overview of our research sites and methods, followed by seven vignettes (two at a repair co-op office and five at an individual repair worker's workshop) that illustrate the range of processes for negotiating repairedness in our setting. We conclude by discussing our contributions to studies of both repair and artifacts in CSCW.

6.2 Conceptualizing "repairedness" in repair work

6.2.1 Repair work as a contingent, ongoing process

This paper draws on CSCW research on repair work, including the influence in this research of STS studies of repair. Repair has long been a central concept in pragmatist and ethnomethodological perspectives within CSCW for framing the situated and materially-embedded processes by which people respond to breakdowns in their work practices (Frohlich et al. 1994; Orr 1996; Suchman 2007). These breakdowns concern not just artifacts, but also procedures, social relationships, or any other structures involved in work practices (e.g., Orr 1996; Henke 1999; Lutters and Ackerman 2002; Jackson et al. 2012).

A major contribution of these studies on repair work has been to question assumptions of order and continuity implicit in many theories of technology design. By adopting a repair lens, these studies instead focus on how the often-assumed "enduring function" of an artifact (Jackson 2017, p.174) is achieved through ongoing repair work rather than being given by its *a priori* design. Artifacts are inherently fragile and mutable; there are myriad ways in which artifacts can break down or decay; and what is broken or what needs to be repaired depends on a particular situation (Büscher et al. 2009; Alby and Zucchermaglio 2009; Denis and Pontille 2015; Dye et al. 2019). Repair is viewed as a contingent, ongoing process that allows us to work and live with artifacts by keeping them usable and useful against shifting practices, environments, and values (Graham and Thrift 2007; Denis and Pontille 2015).

Another emphasis of these studies is to highlight how repair work depend on establishing, mobilizing, and aligning diverse relations of material resources, knowledge, and people (Orr 1996; Castellani et al. 2009; Jackson et al. 2012; Cohn 2016; Huh et al. 2010). Orr (1996), for example, looks at how copy machine technicians at a multinational corporation collaboratively articulate, circulate, and modify their communal repair knowledge in narrative forms, beyond the manuals provided by the corporation. Houston (2019) illustrates how independent mobile phone repair workers in Kampala, Uganda conduct the software testing process called "flashing" by creatively navigating around limited access to manufacturer-provided tools for software tests and repairs.

Studies also point out social and material barriers to mobilizing relations of material resources, knowledge, and people. They discuss problems with "asymmetrical" access to necessary but often-proprietary information between "authorized" and "independent" repair businesses (Jackson et al. 2012). Houston and Jackson (2016) observe how "mechanisms of material closure" such as mobile phone SIM locks as well as infrastructural factors that limit access to parts necessary for repair add constraints to repair workers' operations despite their creative efforts to work around them. Rosner and Ames (2014) illustrate a case of XO laptops distributed to schools in Paraguay, where the broken laptops were unfixable without procuring replacement parts such as screens, which tend to be cost-prohibitive for students and their families in rural areas.

6.2.2 Negotiating repair work: how (much) should an artifact be repaired?

Getting a repair job done depends on negotiating what is expected of repair work — how an artifact under repair is expected to look and function once repaired. Sims and Henke (2012; also Henke 2017; Cohn 2016) argue that repair plays out along a continuum between "repair-as-maintenance" and "repair-as-transformation." Repair-as-maintenance refers to the prioritization of maintaining the *status quo* by restoring original structures and functions, while repair-as-transformation seeks practices to change or reappropriate existing arrangements. Negotiating across these two approaches can create tensions among different expectations and values in repair work (Houston et al. 2016).

Another aspect of negotiation in studies of repair work concerns the process of "letting go" of artifacts as they age (e.g., Tsaknaki et al. 2016). In her study of an aged space science mission infrastructure, Cohn (2016) shows gradual transformations into the end of the life of the infrastructure through ongoing negotiations concerning what features would stay or get discontinued. She argues that the efforts in such projects could be understood as "repair-into-decay" that recognizes "a need to slow down, let go of expectations, and make cuts to functionality" (p.1512). Here, the expectation is not reversing the damages and decays to prolong its intended lifecycle, but a gradual scale-back that takes advantage of the infrastructure's usedness and impermanence.

Negotiating what is expected of repair work is also important since artifacts under repair are unique — artifacts of the same type break down differently and have been used differently by different people for different purposes. Rosner and Ames (2014) illustrate how whether an artifact is worthy of repair depends on how it is actually used in a particular context, and not just its designed or intended use. In their study of urban repair collectives, Houston et al. (2016) show how a repair artifact that was difficult to fix elicited eager participation of volunteers because of the emotional story associated with it.

Finally, negotiating expectations in repair work is also constrained by the materiality of these artifacts. Repair work is viewed as a derivative activity in that the repair artifacts are composed of partially pre-determined parts, materials, mechanisms, and functionalities (Harper 1987; Spelman 2012). Constraints stemming from prior design choices become salient in repairers' attempts to repurpose or reuse the artifacts. Jackson and Kang (2014), for example, find that artistic practices for redefining and recombining found objects were, however creative, limited by the material properties of the artifacts as these artifacts "talk back" (p.451). In their study of the adoption of used PDAs, Huh et al. (2010) similarly observe that, while people had the agency to decide how the devices would be reused or repurposed, some original functionalities and configurations could not be bypassed.

6.2.3 Negotiating a working version of artifacts under repair: when is repair concluded?

The above studies focus on how the processes of performing repair work are contingent, ongoing, and negotiated, but do not give explicit attention to understanding the *concluding*, however temporary, of a repair job. Several studies in CSCW imply that distinctly understanding this process of concluding repair is important. Ludwig et al. (2018) and Pipek and Wulf (2009), in their studies of infrastructural artifacts, show how a repair process ends when designers and users establish a new way of using the infrastructure, comprising a redesigning and reconfiguring of the artifact itself, as well as a realigning of relationships surrounding it. Cohn (2016), mentioned above, shows how ongoing repair processes needed to nonetheless be stabilized at each stage of scaling back the infrastructure over its lifecycle.

These studies indicate that the conclusion of repair work depends on a non-trivial process of negotiating among a multiplicity of versions of an artifact under repair to enact a single version that could be seen as working. The need to negotiate a version of the artifact depends on not just contingently mobilizing resources and relationships for performing the repair work, but contingently enacting *the properties of the artifact itself* that are made visible (Mol 1999; Mol 2002; Law 2002; Suchman 2005). Knorr-Cetina (2005), for example, argues that the artifacts involved in "nonroutine problems" (p.184) — such as repair situations that concern artifacts which break down in many different and unexpected ways — are different from "commodities, instruments, and everyday things" (p.185) in that they are incomplete and unfolding (continuously obtaining and changing their properties), yet simultaneously multiple and singular (taking various forms and instantiations) (pp.191-2). An artifact thus may be viewed as "a pattern of presences and absences" (Law and Singleton 2005, p.342-3) — not as a unitary or static whole but as something that is strategically brought into being by negotiating messy and contingent boundaries of what is present and what is absent.

While an artifact entails multiple versions that draw on a contingent set of properties, our engagement with the artifact at a given moment in a given situation is particular and singular (Redström and Wilste 2018) — that is, the properties of the artifact need to be made to "hang together" as a singular object (Mol 2002, p.5; Law 2002). The conclusion of a repair job critically depends on negotiating one working version out of multiple possible versions of an artifact under repair by differently configuring its properties according to the repair situation. Along these lines, our study seeks to unpack the processes of negotiating a working state of an artifact under repair — or what we refer to as the artifact's *repairedness*.

6.3 Methods and analysis

6.3.1 Field sites

6.3.1.1 The repair community at the Sewoon Electronics Market

Established in 1968 in Seoul, the Sewoon Electronics Market has long enjoyed status as Korea's biggest and oldest home for analog electronics repair experts. Deterioration in recent decades, as well as the rise of the digital-focused Yongsan Electronics Market, led successive mayors to suggest its demolition. In 2014, the Seoul Metropolitan Government decided instead to renovate the structure as part of the city's regeneration plans. A key initiative of the revamp was to evolve Sewoon into a new hub for urban manufacturing by facilitating collaborations between experienced repair workers and the newcomer group of tech startups and makers. The first phase of the renovation was completed in 2017 with the redesign of the building's façade and the addition of new spaces for tech startups, makers, and artists. Our study took place within this collaborative context, in that we sought to better understand and articulate the repair workers' processes for getting their jobs done.

Our ethnographic study of repairers at Sewoon offered rich insights into how people engage with the properties of artifacts through their processes of performing and concluding repair work. Repair workers at Sewoon mostly were veterans with decades of experience who dealt with a wide range of electronics, lab equipment, and industrial machinery. The artifacts brought into Sewoon in many cases required major effort and expertise just to figure out their structures and functions: many were custom-made and their schematics were unavailable; others arrived as parts (e.g., circuit boards) extracted from the whole of a piece of machinery due to portability issues. Due to the idiosyncrasy and age of many of the artifacts, replacement parts were often difficult to source.

The artifact's owner and the repair workers often noticed different problems with different properties in the same artifact, which translated into different expectations of what would make the broken artifact work again and reach a state of repairedness. This study mainly draws on fieldwork at two specific sites — a repair co-op and an individual repair workshop — within Sewoon to analyze the processes for achieving repairedness in repair work. All of the fieldwork was conducted by the first author.

6.3.1.2 The repair co-op

The fieldwork began by shadowing the manager, JIN, of a two-year-old repair co-op at Sewoon. With the tagline "We repair your memories," the co-op started as a pilot project to help revitalize the Sewoon community. It involved six repair veterans with different specializations, all of whom were men in their sixties and seventies. The main office of the co-op was also the personal workshop of SUNG, the co-op president. JIN, a female in her thirties, acted as SUNG's assistant. Artifacts in various repair stages, some of them partially disassembled, were spread around the office, and those that needed to be repaired or had been donated by the customers after being declared "unrepairable" were stacked on the shelves with post-it notes such as "re-repair" and "for parts salvaging." The other members of the co-op often visited the office for chats over instant coffee (sachets of pre-mixed instant coffee, cream, and sugar) drunk from disposable paper cups filled from the hot water tap of a water purifier. This "mix coffee" was ubiquitously enjoyed at Sewoon, and serves in Korean work environments more generally as the equivalent of the proverbial water cooler.

The co-op had a website for managing repair requests, but a majority of their customers visited the co-op office in person with their artifacts. The manager JIN reviewed the forms or the types of artifacts that came in to filter out requests that fell outside the members' areas of expertise. Once accepted, she entered the customer's contact details, the type of the artifact, and symptom(s) provided by the customer into a digital file on the desktop computer in the office. The co-op made it clear that the cost and time required for repair, or the repairability itself, could only be determined after a member of the co-op had examined the item.

When an artifact arrived at their office, the manager JIN discussed with the co-op president SUNG which member would be matched with the artifact based on the co-op members' diverse experience and expertise. After selecting a member, JIN would make a courtesy call to the matched member to notify him of the new item, and then she would physically deliver it to his workshop at Sewoon. While the selected member would often complain ("I don't have time for this," "this kind of stuff is tricky to fix") or ask if there were no other members who could take on the job, they rarely rejected a job without examining the artifact themselves first. During the year prior to our study, the co-op received around 800 items to repair. 88 percent of these items had been repaired ("repair completed"), with the rest recorded as "repair on hold." In addition to helping allocate incoming artifacts to suitable members, the manager JIN performed several roles, such as dealing with logistical and communication problems among members and between customers and members; testing artifacts once they were repaired; and arranging shipping.

Prior to conducting fieldwork at the co-op office, the first author visited the office and provided JIN and SUNG with an overview of the proposed study as well as what the methods would entail (e.g., observation at and around the office, interacting with other co-op members, and documentation such as photographing, audio-/video-recording, and note-taking). After obtaining their consent to participation, interactions and events that happened throughout the manager's day were observed and documented two times a week for two months. Once the first author got acquainted with the other members of the co-op beyond JIN and SUNG, she conducted semi-structured, in-depth interviews with four of them [RC3-6; see **Table 5**] in their individual workshops. The interviews usually lasted around an hour, revolving around challenges they faced in doing repair work, how they dealt with these challenges, and how they managed their personal inventories of parts and tools. Participants gave guided tours of their workshops and parts inventories. With their permission, their actual work processes were observed, photographed, or video-recorded where possible. While the fieldwork at the repair co-op and interviews with the co-op members involved no financial compensation, the first author occasionally offered the participants coffee or juice as a token of appreciation.

6.3.1.3 CHO's repair workshop

About a month into our fieldwork at the co-op office, the first author was introduced to CHO, who was not a co-op member. When the first author entered his workshop on the sixth floor, CHO greeted her sitting in front of thousands of labeled plastic drawers that filled up the two sides of the space. Around the shop were additional drawer units, piles of boxes, and vintage machines.

CHO had over 20 years of repair experience, specializing in electric pipe organs and industrial and medical equipment. He was known in the community for his theoretical knowledge and extensive inventory of electronic parts, which he cataloged in detail in online spreadsheets. His spreadsheet for transistors alone, for example, included around 1,400 types of transistors whose total number of units in stock neared 190,000. The repair workers at Sewoon often came in to ask for his help in identifying damaged parts on circuit boards, sourcing replacement parts, or finding cross-references to parts and their corresponding online datasheets. Similar to the other repair workers at Sewoon, CHO engaged with multiple repair projects asynchronously. Many of the artifacts at his shop came from his long-term customers, including research institutes and labs. A self-described “slow repairer,” he preferred to work on these industrial and specialized items than consumer electronics. They were more “unique and interesting” while giving him more control over the repair timeline, which varied between a few hours and years. Even when he failed to repair a machine, it often just stayed with him as the customer could not find anyone else to fix it. Over time, these unrepaired or unrepairable machines accumulated in his space.

During the first meeting, CHO was briefed on what the proposed research was about and what the methods would entail. CHO agreed to participate in the study and suggested that the first author assist him at his shop to better observe his work practices. He insisted that he pay for her assistance so that he would feel comfortable working with her, however informally. They agreed on weekly payments of 50,000 Korean won (less than 50 US dollars) during the fieldwork.

The first author worked at his workshop three days a week for three months, assisting him with organizing and managing his inventory both online and offline, researching schematics, delivering parts and devices to other repair workers that he worked with, greeting customers, invoicing, shipping repaired devices, and simple repair work such as assembly and soldering. He provided the first author with a desk and a PC whose monitor was connected to his PC on his main workbench so that they could work on the same document simultaneously. The first author also observed his repair processes by photographing or video-recording them, along with spontaneous, think-aloud, and in-depth interviews. CHO invited the first author into his online group chat rooms that he used to communicate with his customers, in which documents and photographs relating to respective repair projects were shared. The chat room participants acknowledged her as CHO’s assistant. The first author observed the interactions in these chat rooms both during and after the fieldwork, making the total duration of the chat room observations 18 months.

Table 5. List of participants

Code [pseudonym]	Gender	Age	Occupation$^{\alpha}$ [area(s) of expertise]	Location within Sewoon	Participation type [vignette(s) (section)]
Site 1. Repair co-op					
RC1 [JIN]	Female	30s	Repair co-op manager	Co-op office (2F)	Observed & interviewed / shadowed by the first author [**No bass** (6.4.1.1) and **Low volume** (6.4.1.2)]
RC2 [SUNG]	Male	70s	Repair co-op president / repair worker [vintage audio equipment]	Co-op office (2F)	Observed & interviewed [**No bass** (6.4.1.1) and **Low volume** (6.4.1.2)]
RC3 [YOON]	Male	60s	Repair co-op member / repair worker [audio equipment & electronics]	Individual shop (5F)	Interviewed [**Low volume** (6.4.1.2)]
RC4	Male	60s	Repair co-op member / repair worker [audio equipment & electronics]	Individual shop (5F)	Interviewed
RC5	Male	60s	Repair co-op member / repair worker [home appliances & electronics]	Individual shop (4F)	Interviewed
RC6	Male	70s	Repair co-op member / repair worker [circuit design & power electronics]	Individual shop (4F)	Interviewed
Site 2. CHO's workshop (the pipe organ project)					
PO1 [CHO]	Male	60s	Repair worker [pipe organs, industrial equipment, circuit design & electronics]	Individual shop (6F)	Observed & interviewed / shadowed and assisted by the first author [**Pipe organ** (6.4.2.1-6.4.2.5)]
PO2 [NAM]	Male	60s	Repair worker [electronics – general]	Individual shop (6F)	Interviewed [**Pipe organ** (6.4.2.2)]
PO3 [LEE]	Male	60s	Repair worker [circuit design & integrated circuits]	Individual shop (5F)	Interviewed [**Pipe organ** (6.4.2.2, 6.4.2.3 & 6.4.2.5)]
PO4 [KIM]	Male	50s	Administrative staff	Church	[**Pipe organ** (6.4.2.1-6.4.2.5)]
PO5 [PARK]	Male	50s	Director	Church	[**Pipe organ** (6.4.2.5)]
Other participants at Sewoon					
SR1	Male	70s	Repair worker [circuit design & integrated circuits]	Individual shop (3F)	Interviewed
SR2	Male	50s	Repair worker [circuit design & integrated circuits]	Individual shop (3F)	Interviewed

SR3	Male	70s	Repair worker [industrial & medical equipment]	Individual shop (4F)	Interviewed
SR4	Male	60s	Repair worker [motors]	Shared shop (4F)	Interviewed
SR5	Male	50s	Repair worker [motors]		Interviewed
SR6	Male	60s	Repair worker [electronics – general]	Individual shop (5F)	Interviewed
SR7	Male	70s	Repair worker [vintage audio equipment]	Individual shop (7F)	Interviewed
SE1	Male	70s	Engineer/developer [circuit design, audio equipment & sensors]	Individual shop (8F)	Interviewed
SE2	Male	50s	Engineer/developer [circuit design & LED control]	Individual shop (3F)	Interviewed
SE3	Male	40s	Engineer/developer [3D printing]	Individual shop (3F)	Interviewed
SP1	Female	40s	Educational electronic kits manufacturer / supplier	Shared shop (5F)	Interviewed
SP2	Male	50s	Electronic parts supplier	Individual shop (3F)	Interviewed
SP3	Male	40s	Electronic parts supplier	Individual shop (3F)	Interviewed

[a] The titles "repair worker" and "engineer/developer" were often used interchangeably at Sewoon. Participants were listed as either "repair worker" or "engineer/developer" depending on how they wanted to be identified.

6.3.2 Data analysis

The fieldwork produced interviews, photographs (>1,200), and videos (>52 hours) of repair workers' inventories and work processes, online chat room logs as well as extensive field notes. Photographs, including those containing the participants' identifiable features, were taken and included in this paper with the permission of the participants. The interviews and chat logs were transcribed and collected in the participants' original language (Korean), and were translated into English by the first author, who is a native Korean speaker. In translating technical terms, the first author consulted with CHO, who was experienced in working with technical documents written in English. In addition to the participants at and around the repair co-op and CHO's workshop, the first author conducted semi-structured, in-depth interviews with seven repair workers, three engineers/developers as well as three parts suppliers at their individual shops to gain a bigger picture of repair work at Sewoon. These interviews lasted between 30 minutes and one hour, focusing on challenges associated with repair work, parts supply, or personal repair inventory management. Participants were given code names and pseudonyms, if they appeared in the vignettes included in this paper (see **Table 5**). The transcripts have been lightly edited to improve readability.

Data collection happened in three rounds as the fieldwork sites and research themes changed over time. The first month of data collection was geared towards getting a sense of the repair community and the artifacts being repaired at Sewoon, involving data from the initial fieldwork at the repair co-op as well as a small set of exploratory interviews with repair workers outside the co-op. The first author transcribed and open-coded interviews, photographs, videos, and field notes to identify potential themes. These themes, including repair inventory management, and formal and informal partnerships surrounding a repair project, were discussed with the other authors to inform the next round of data collection. Data collection in the second phase involved data gathered from fieldwork both at the repair co-op and CHO's workshop. This data was also analyzed and discussed with the other authors, from which a theme of negotiations surrounding an artifact under repair emerged. The last round of data collection happened mostly at CHO's workshop. Additional interviews with repair workers and parts suppliers at Sewoon were conducted to confirm the findings from the previous rounds of analysis.

Given the importance of examining both repair workers and artifacts in particular repair situations to our study, we adopted situational analysis (Clarke et al. 2015) to analyze our data. Situational analysis is an updated version of grounded theory that, among other advances, attends to the complex ecology of relations (Star and Ruhleder 1996) among both human and non-human actors within and surrounding the situation of inquiry. Using this approach, we first laid out what appeared to be the key human and non-human actors (e.g., repair workers, customers, repair parts and tools, artifacts being repaired) and sociocultural elements (e.g., repair rates, co-op membership, informal partnerships among repair workers, customers' and repair workers' expectations of *working* artifacts, Sewoon as the last resort for analog electronics repair) in our setting, and explored relationships among these elements. We then examined how these relationships surfaced and were negotiated in particular repair situations as well as the positions that actors took or did not take. Through the iteration of this process, the need to negotiate multiple versions of artifacts under repair emerged as a key theme, and questions and focuses were modified accordingly. We also regularly consulted with the key informants (the co-op manager JIN and CHO) to communicate and confirm our tentative findings.

Despite our focus on introducing and framing a concept of repairedness to characterize the conclusion of repair work, we were limited in our collection of data to onsite interactions, and not to the owners' detailed accounts of their artifacts or the outcomes of repair work. Because of our agreements with the repair workers, we did not contact the owners separately to collect data. In those cases where the owners were present at the site, they were never interviewed separate from the repairers so as not to damage rapport with the repair workers. In contrast, the repairer-owner interactions during the repair process occurred mostly over the phone or in the chat rooms, which could be observed and which we included in our data. The repair workers were also asked during interviews to describe in detail their interactions with owners of artifacts under repair, allowing us to grasp a range of interactions and the role of the owners in repair projects and their conclusion.

6.4 Results: negotiating repairedness

To characterize the process of negotiating the repairedness of artifacts at Sewoon, we present seven vignettes. In the first vignette, a repair worker negotiated the repairedness of a record player and vintage amplifier differently according to whether addressing their identified properties (i.e., static noise, bass sound, a working motor) was straightforward. In the second vignette, negotiating the repairedness of a radio also required determining sufficient levels of certain properties, specifically volume. In the last series of five vignettes, a repair worker negotiated the repairedness of an electric pipe organ with the members of a church differently over time as the priority and sufficiency of the properties that were identified shifted during the process of performing the repair work.

6.4.1 At the repair co-op

The following two vignettes at the repair co-op illustrate how identifying the properties of an artifact relevant to its repairedness was central to concluding repair work. The key point is that negotiating properties may be needed since the notion of a working artifact can mean different versions of the artifact to different people, corresponding to different prioritizations and sufficient states of these properties. Repairedness was not necessarily an objective result of repair work whose path could be charted out in advance, but rather was a contingent result of the repair worker and the owner negotiating a version of the artifact. The vignettes were reconstructed from the first author's field notes.

6.4.1.1 The Braun record player and the no-bass amplifier

One afternoon, SUNG, the co-op president, finished working on the repair of a vintage Braun record player from the 60s. As SUNG placed a vinyl record on the record player, classical music blared out from the speakers on the rack above his workbench. Until recently, he had deemed the machine "unrepairable" since it needed a rare replacement motor that he didn't have. He had put it aside for six months. One of his customers recently happened to donate the same machine to the co-op, which was broken beyond what the customer was willing to invest in to repair but which still had a working motor. Using that motor as the replacement, the Braun record player was now working again and blaring music.

After a short break, he moved the repaired record player to another table and grabbed a partially disassembled vintage amplifier that SUNG deemed still in need of repair. The owner had wanted its static noise problem to be fixed, which SUNG had already done by replacing some vacuum tubes. SUNG had found, however, that the amplifier also produced no bass sound and thus was still in need of repair. SUNG stared at the amplifier for a while and unplugged some vacuum tubes from its circuit board and inspected them. He then took out a thick faded vacuum tube data book and flipped through it, removing and reseating the tubes for a while with no luck.

"This McIntosh C1, can you find the schematic?" SUNG asked the co-op manager JIN, staring into the amplifier.

"McIntosh C1, you said?" JIN grabbed a chair and sat in front of the desktop computer. She searched online and found no schematics for a McIntosh C1.

"Is that really a McIntosh C1?" she asked.

He did not answer, examining the board. JIN kept browsing through the online search results. Silence continued for a while.

"Do you have his [the owner's] number?" SUNG asked, breaking the silence. She opened the customer database on the computer.

"Hmmm. He did say it's a C1. Here's his number," she highlighted the number on the screen for him to see. SUNG made a call.

"Hi, repair shop here. So about your McIntosh C1...We've replaced all the parts and fixed the static noise, but there's almost no bass...I think there's a problem with the circuit...The volume's fine but there's no bass sound...It could be tricky, I'll need to take a closer look at it...OK, got it," he ended the call.

JIN, who overheard the conversation, explained to me that SUNG told the owner about the new problem (no bass sound) but apparently the owner was happy that the static noise problem had been fixed and did not ask for further repair.

"So the repair is done?" I asked her.

"Yeah, it's done."

Analysis: negotiating properties of repairedness as they emerge. The two artifacts of the Braun record player and vintage amplifier in the vignette highlight two extremes of this contingent process of identifying properties to negotiate repairedness. SUNG could fix the Braun record player by identifying the property of a working motor that initially could not be replaced, such that the record player could not reach a state of repairedness. When he tested the record player after replacing the motor, the music played with no apparent problems so SUNG simply concluded the repair. In contrast, when SUNG went to fix the static noise of the vintage amplifier, repairedness was not achieved, as he detected a problem regarding another property of "no bass sound." In SUNG's version of the artifact, the property of bass sound was an additional condition for the amplifier to be considered working.

Achieving repairedness in SUNG's version depended on identifying an initially unknown set of additional properties that would be relevant to resolving the no-bass sound. To do so, he tried to pinpoint the source of the remaining bass problem by first inspecting vacuum tubes and connected parts on the circuit board of the amplifier. When this inspection did not yield actionable clues, he considered inspecting the amplifier's underlying connections among onboard components against a schematic of the machine. Inspecting at this level could be "tricky," requiring more time and effort to analyze and test a broader range of properties of the amplifier, not to mention that the schematic was not readily available. SUNG needed to discuss with the owner what should be done about the bass sound.

For the owner, the key properties for considering his amplifier to be working again were the absence of static noise and, we presume, an appropriate level of the output volume ("the volume's fine"). The first condition had become visible with its breakdown. The second property only emerged in relation to SUNG's identification of the no bass problem. The owner, however, viewed the bass sound property as less critical, without which the amplifier could still work *for him*. When the key conditions were prioritized and restored, the repair worker and the owner agreed that the repair of the amplifier could be seen as concluded. Repairedness was thus the negotiated result of identifying two versions of the artifact (the repair person's and the owner's) and settling on one (the owner's).

The variation in the effort for identification needed to negotiate the repairedness of the Braun record player versus the vintage amplifier shows how repairedness is achieved differently for artifacts being repaired according to specific owners and repair workers. In the case of the Braun record player, a working motor was a non-negotiable property for both SUNG and the owner, without which the record player could not be considered repaired. Alternative versions were not considered, and thus identifying and prioritizing properties to achieve repairedness was straightforward. In the amplifier case, the repairer worker and the owner could consider multiple possible versions of the artifact as a working amplifier based on what properties needed to be identified and prioritized.

6.4.1.2 The low-volume radio

A few days later, JIN was about to bring a turntable that just came in that day from the co-op office to the workshop of YOON, a co-op member in his late 60s whose expertise covered a vast range of analog electronics. JIN headed to YOON's shop. YOON was not at his shop and the door was locked.

"He must be on a cigarette break," she said. JIN rang YOON. Still on the phone, she handed the turntable to me. She found keys on the top of the doorframe and unlocked the door.

"The small one next to R3000? Got it, I'll take it," she said to YOON.

Hanging up the phone, she put down the turntable on his workbench. Next to the turntable on the workbench, she grabbed a Nivico radio with a sticky note that read "12/28." She told me that YOON said that it's "partially repaired."

We returned to the co-op office, bringing the Nivico radio. When we got inside, she turned the radio on. The radio didn't seem to be working. She switched it off and on again.

"It doesn't seem like this has been fixed at all. Maybe I took the wrong one."

We headed back to YOON's shop, where YOON had now come back from his cigarette break.

"Is this the right one? It doesn't play at all," she asked.

YOON snatched it from her and adjusted its antenna. The radio did play, but the volume was pretty low.

"This is as loud as it gets. If the owner really wants to have it fixed, I'll need to replace all the parts, but is it worth it? I don't know," YOON questioned. "I know they feel attached to these (machines), but that comes with a price. Are they really willing to pay for it?"

He handed the radio back to JIN and we came back to the co-op office.

JIN reported to her boss SUNG: “YOON says this is as loud as it gets.”

“What good is it then?” retorted SUNG.

JIN went to the PC in the office and looked up the owner’s contact number and information about the radio from the database. According to the database, the vintage radio had been brought in with a power problem — it would not switch on. She then gave a call to the owner.

“Hi, repair shop here. Good, good. So we have partially fixed your radio. It plays, but the volume’s really low...We don’t know what’s wrong. If you want to fix that (volume problem), we can look into that...Would that still work for you?...Right...OK, we’ll ship it out today.”

Hanging up, she told SUNG: “Well, she says that [the radio with low sound levels] works for her.”

Shrugging, he returned attention to an audio system that he had been working on.

Analysis: negotiating sufficient states of properties of repairedness. The first vignette illustrated how the properties for determining the repairedness of the record player and amplifier were not given, but needed to be identified and the resulting versions of the artifact negotiated. In this second vignette, the repairedness of the Nivico radio was not just about fully restoring the properties that were identified but about negotiating the sufficient states of certain properties, contingent on the particular repair situation. The repair situation, for example, mediated the sufficiency of the Nivico radio’s volume level for YOON and the owner.

YOON believed the radio was “partially repaired” but that, given the trickiness of a full repair, it could nonetheless be considered in a state of repairedness. YOON did not see the low volumes as problematic enough to make the radio remain unusable for the owner. Fully restoring the volume would require additional labor, which might not be compensated for. When JIN notified the owner of the new problem with the sufficiency of the volume and potential complications for addressing the problem, the owner decided to conclude the repair as the artifact was seen as sufficiently working as a radio for her. We may speculate that she did not want to prolong the repair for a potentially not-so-dramatic improvement given that she had submitted the radio to the co-op more than five months ago. Regardless of her motivation, the different situations of YOON and the owner needed to be resolved to negotiate a version of the artifact with a sufficient volume level that would achieve repairedness.

6.4.2 CHO's project of repairing a church's electric pipe organ

The following five vignettes illustrate how the repair worker CHO negotiated with the staff of a church to achieve a state of repairedness for the church's electric pipe organ. As in the vignettes at the co-op office, achieving a state of repairedness involved identifying, prioritizing, and evaluating the sufficiency of a set of properties. The key insight here is that the priority and sufficiency of the properties involved in the negotiation of repairedness also *shifted* throughout the organ's repair process.

CHO's desire to get paid in a timely manner and limit the time that he allocated to the repair work induced the first shift — CHO's negotiation of a more *limited scope* of repairedness with the church staff. Achieving repairedness also involved dealing with the structural and functional multiplicity of the organ, as well as problems with specific properties that unpredictably recurred or emerged during the repair process. This led to the second shift, in which properties that were initially seen as "problematic" became negotiated as "working but still under repair," and then "sufficiently working" even without a full repair. The vignettes were reconstructed from the first author's field notes, interviews, as well as the texts, videos, and photos shared on an online group chat room and an online photo album.

6.4.2.1 Making repair scenarios

CHO got a call during our lunch from one of the biggest Korean Methodist churches in the southern part of Seoul, about a one-hour drive from CHO's office. He had helped to install an electric pipe organ for the church in the mid-1990s. The church's administrative staff explained the basic problem: it was taking too long for the organ to boot up, which interfered with their tightly-scheduled church services. They asked CHO to come down to the church and have a look. CHO did not seem eager to take on the project and told the staff that he would call back. He explained to me that the fact that he had installed it did not mean that he could repair it. He anticipated a cascade of problems that could be difficult to fix. Since he could not carry all repair tools that might become necessary onsite, he would need to extract and transport parts and tools multiple times. Despite all the uncertainties, the church would not pay until the problem was resolved.

CHO sat down and wrote out two scenarios for resolving the boot up problem to communicate with the church staff regarding billing conditions: (i) *"DC power supply repair"* If no particular systemic problem is identified on the first visit, all of the circuit boards would have to be extracted and then reinstalled after replacing any aged or problematic components. If no further problem were identified within two weeks after reinstallation, he could then consider the organ repaired and charge accordingly; (ii) *"System error repair"* In case that problems were to persist, then he would have to additionally analyze the organ's overall electrical system and perform further repairs to address these problems. These repairs would incur additional costs, and the warranty would last six months after the completion of this round of repair. He emailed this document to the church staff for confirmation. The staff agreed to the billing conditions over the phone.

Analysis: limiting the scope of repairedness. In this first phase, the church staff believed that the organ only had a problem with a single property of slow boot times. Repairedness for the staff concerned making the organ switch on immediately or at least sufficiently fast. CHO knew from his experience that achieving this seemingly simple state could be more complicated: (i) the boot time issue could be linked to multiple properties of the organ that all needed to be repaired; (ii) working on extracted parts would add complexity in testing those properties; and (iii) the problem might recur after the first round of repair. The payment, however, would only be made after the repair had been concluded.

To identify and negotiate a limited scope for repairedness of the organ and make the repair project manageable, CHO identified two scenarios (power supply repair and system error repair), each focusing on a subset of properties versus the overall set of properties that he knew from experience might be relevant to the boot time problem. For the repair to officially begin, the church needed to agree on CHO's scenarios, that is, what properties would be inspected and how long a repair would be subject to warranty. While the actual repair process might diverge from these scenarios, negotiating the scope of the repairedness of the organ was necessary not only for the conclusion of the repair work, but even for its initiation.

6.4.2.2 Connecting the organ's properties across the console and the control room

Two days later, CHO, NAM (CHO's fellow repair worker), and myself arrived at the church. The church comprised two skybridge-linked glass buildings, a four-floor building that housed the auditorium and the later-built 16-floor building used for bible classes, administration offices, and guest lodging. As we entered the auditorium, a church staff member, KIM, greeted us standing next to the organ. The organ was on the left side of the auditorium stage, taking up a portion of the wall behind the stage. This particular organ was designed in the mid-1990s by a Korean sound engineer trained in Germany. CHO had been responsible for installing its electric circuits using prefab parts, which, according to CHO, had been expected to last at least 40-50 years in ideal conditions.

CHO quickly scanned through the organ's console, which had three sets of keyboards at the center, two small LED screens above the keyboards, four rows of labeled stop knobs that control ranks of pipes on both sides, and pedalboards below the keyboards. Beneath the keyboards were also a row of black "memory" buttons to program and switch around different play settings. On the right side were two small white buttons for the lighting and the power. CHO asked NAM to stay at the console for testing.

I followed CHO to the organ's control room behind the auditorium. With the assistance of the church admin staff, we unlocked and opened the massive wooden door. Rows of metal pipes filled one side of the dark and dusty room. We switched on the light and found a system rack in a plastic case facing the pipes. CHO opened the plastic case, creating clouds of dust. We started coughing. Once the case was removed, he extracted each of the ten circuit boards from the rack to look for any visible signs of failure. He found none. He kept the CPU board and asked me to reinsert the rest tightly. He took out two pieces of testing equipment, which he connected to the board that he kept.

He then shouted through the control room wall: "NAM, can you turn on the organ?"

As we heard rumbling sounds, the number on one of the testers read 5.06V. The organ was on, electrically speaking, though not as a musical instrument.

"Can you test the organ? Press memory buttons and you'll see numbers on the LED screen. See if you can play anything," CHO asked NAM.

"What do you mean by memory?" NAM yelled back.

"Ah, you don't know how to play it? Pull a few stop knobs, the black stick-like ones, and see if that makes any sounds...No? Not yet?"

NAM pressed the keyboards and pedals, but no musical sounds came out.

"So no sound, huh? Switch it off, then," CHO shouted. The voltage rating on the screen dropped to zero.

In about a minute, CHO asked NAM to turn on the organ again. NAM observed no responses. About 50 seconds on, we suddenly heard a burst of organ sounds.

"It works now!" NAM yelled from the auditorium.

After testing the organ's stop knobs, display, and memory settings, CHO asked NAM to switch off the organ. When NAM switched on the organ again, it started working immediately. CHO reasoned that the slow boot time could have been caused by some problems in the circuit boards, as he could not read any numbers particularly off from the testing equipment. For more detailed tests, we extracted and boxed the organ's ten circuit boards and headed back to Sewoon.

Back at CHO's office, CHO sent the whole box of circuit boards to LEE, his frequent collaborator at Sewoon, simply saying that they were for an electric pipe organ. LEE, who specialized in industrial circuit board design and repair, would do testing and get his share when the church made a payment. In the late afternoon, LEE called CHO to say that there seemed to be a problem with the CPU board's reset function. Based on LEE's description, CHO came up with an explanation that one of the capacitors' capacity decay had caused the reset malfunction. CHO asked LEE to replace the potentially problematic capacitor. Once CHO got back the boards from LEE, he tested the CPU board around the replaced capacitor. It seemed to be working fine.

CHO and I went back to the church the next day with the tested circuit boards. As the church staff watched, I reinserted the boards into the rack in the organ's control room, and then we headed to the organ's console in the auditorium. CHO pressed the power button. After a few seconds, the LED screen lit up. He then switched it off and on again. The screen lit up immediately with a quick rumbling sound, and everything seemed to work sufficiently. The admin staff KIM was satisfied with the result. CHO and KIM agreed on monitoring the organ, without any further repair needed.

Analysis: connecting properties of repairedness. This vignette illustrates how the structural and functional multiplicity of the organ mattered in that achieving repairedness required making connections between physically separate properties. Properties of the organ relevant to the initial focus on the boot time issue were distributed across a complex web of interconnected parts and functionalities within and between the console and the control room. The console's physical operation involved LED screens, keyboards, pedalboards, stop knobs, memory buttons, and the power button. The operation of these parts was connected to the ten circuit boards and the electrical system in the separate control room.

While CHO speculated that the properties connected to the boot time could be found in the CPU board, he could not test or repair the board onsite without the necessary testing equipment. The whole ten boards were extracted and transported to the repair workshops at Sewoon. The potentially relevant property (the capacitor on the CPU board) was identified through testing and repaired, and the boards were reinstalled into the system rack in the control room to check the organ's overall operation. The boot time problem was resolved, and the organ seemed to have achieved a state of repairedness.

Different properties of the organ had different levels of visibility to different people, leading to different expectations of repairedness. For the church, the organ was equivalent to the musical instrument in the auditorium: it should boot up reasonably quickly, play as programmed and without any electrical interruption. For LEE, the organ was almost reducible to its circuit boards and electrical mechanisms: it worked when the boards and their components were free from electrical problems. CHO similarly saw the organ fundamentally as an electric machine whose operation was programmed in and controlled by the circuits. However, he also needed to build, translate, test connections between what happened in the control room and what manifested in the console to identify properties that mattered for repairedness and relations among those properties. For him, the organ became repaired when the particular properties in the console, the circuits, and the electricity system were aligned and repaired.

6.4.2.3 The persisting problem in light of new problems

KIM called CHO the next morning and said the organ was now “completely problematic” and that they would need immediate attention to it since they were to have a big service scheduled later that day. He said it would take 34 seconds for the organ to boot up and that the organ would also shut down randomly. CHO and I hurried to the church. CHO went to the switchboard in the control room and found that the circuit breaker had tripped. CHO repaired the circuit breaker, and the power was restored. In addition to the power problem, the church’s organist complained to CHO that its memory settings had been wiped out. She pushed the memory buttons, each of which corresponded to a different play setting, but the stop knobs did not move in and out properly. “It’s just not working,” she said. After some inspection, CHO found that the stops labeled between 33 and 48 did not work. He left a note on the organ saying: “Please operate manually the stops from 33 to 48 (memory problem).” We took out the circuit boards again to see what more could be done.

Back at Sewoon, CHO asked LEE to look at the boards again in relation to the memory problem. LEE could not identify any failures. CHO took over and studied the CPU board, and concluded it could just be that the fuse holder became loose while in transit. A few days later, we returned to the church with four pieces of testing equipment and the boards now wrapped in bubble wrap to protect the fuse holder. This time, KIM followed us throughout the repair and testing process, taking pictures with his cellphone to record the repair process for their internal reports. The boards were reinstalled. It now took 42 seconds for the organ to start, but the memory problem seemed to have been fixed. CHO and KIM agreed to monitor the organ for the time being.

It took only two days for us to get another nervous call from KIM that the organ would not boot up for at least seven minutes now. They had another service scheduled within a day. CHO went back to the church and came back with the extracted boards again. CHO and LEE performed additional tests on the boards but detected no obvious problems. The boards needed to be reinstalled regardless for the organ to be used for the church's regular services. This time, the organ turned on in about a minute. CHO did not know what caused the variance in boot times. He suggested that the organists just keep the organ on while his analysis continued. The admin staff reluctantly agreed and added a note to the organ that read: "For the time being, please do not switch off the pipe organ on Saturdays and Sundays (repair in progress)."

Analysis: re-prioritizing properties of repairedness. In this vignette, we highlight how a single repair job also may involve multiple processes that arise as repair problems recur or emerge in unpredictable ways. The key point is that the recurrence or emergence of new repair problems is mediated at every step by the need to negotiate new versions of the artifact, such that the priority of the properties of repairedness is continually shifting.

In the organ project, the boot time problem recurred while new problems with the organ that had to do with different properties also emerged. The church staff initially perceived the issue with the boot time property of the organ to mean it was almost in a state of repairedness, but this view shifted in light of the new problems of memory failure and power failure. As these problems emerged, the church staff and the organist viewed the organ as now "completely problematic" and "just not working." The memory and power properties therefore shifted to be prioritized over the persisting boot time problem. The organ, for it to be considered repaired, now needed to be free of these additional problems. While these problems were then resolved, the boot time issue remained unaddressed and CHO could not figure out what caused the issue. The church, however, now agreed that the boot time was an issue subject to monitoring rather than immediate action. The properties relevant to repairedness thus shifted as the priorities shifted.

6.4.2.4 Working again but still under repair

CHO continued to extract, test, and reinstall the circuit boards but identifying the boot time problem was elusive. The nervous church staff, meanwhile, kept the organ on during the weekends. The circuit boards seemed to work well when tested at CHO's and LEE's offices. On CHO's next visit to the church, however, they were reinstalled into the organ's system and did not work properly. CHO explained that it could have been much easier if he could work with the circuit boards while the organ was in operation. "This is like being a house call doctor, which I am not. I'd like to bring my patient to my operating table to perform surgery, where I'll have all the tools I want," he quipped.

CHO performed what had become a routine test over the next several weeks. Potential problems were detected, including some noise with the CPU, the CPU board's contact failure, and the ripples around the rectifier. Additional tests, however, failed to yield actionable properties. As what caused the boot time problem was still unknown, CHO decided on a makeshift solution: he placed a benchtop power supply in the control room and connected it to the CPU board. This way, the board would continue to be powered even when the organ's console was switched off, allowing the organ to switch on immediately. CHO was going to build a more stable and portable backup power supply to replace the benchtop one. Over the next several months, the church admin staff KIM occasionally inquired when it would be installed. These check-in messages usually read as follows: "The organ's working without trouble. When will the repair be resumed?" With no apparent problems with the organ, and with other projects that CHO needed to work on, the backup power supply installation was delayed.

Analysis: temporarily substituting properties of repairedness. The need to extract the circuit boards and test them outside of their original context of operation created additional uncertainties, compared to our previous vignettes from the repair co-op where the repairers could engage with the whole artifacts under repair in their workshops. The extracted boards tested in the workshops and the boards installed in the organ's system rack operated differently, and the properties of the same boards in multiple contexts should somehow be aligned for CHO to work towards a state of repairedness.

The drawn-out process of identifying technical properties related to the boot time made acceptable their temporary substitution with the property of *staying powered*, both from the perspective of the organ users from the church and from CHO's perspective as a repair worker. The church needed to use the organ for their regular services. Not switching off the organ and later installing a temporary power supply became acceptable stopgap solutions for the church staff as these solutions removed the practical issue with the boot time property by making the organ switch on immediately. For CHO, the boot time property identified by the church staff also included the technical properties — namely, the electrical properties of the circuit board, which were yet to be identified or addressed. The stopgap solution also worked for CHO, however, as he wanted to conclude the repair project and get paid. Without having been fully repaired (the underlying electrical properties had not been fully resolved), the organ reached a state of repairedness, where the involved parties acknowledged the organ was still under repair electronically but working again from the practical point of view of the organ users.

6.4.2.5 "Fully" repaired, for now

CHO now wanted to conclude the repair with the installation of the backup power supply so that he could finally get paid for his work. The church's admin staff KIM and the church director PARK agreed with his plan. The board members of the church, who were being kept regularly informed about the organ's repair status by the admin staff, were soon to be replaced with new members, and this would complicate the billing process. On what was expected to be the final visit, CHO disconnected the benchtop power supply from the CPU board and installed the custom-built backup power supply. The organ did not turn on for minutes. He once again extracted three boards responsible for the power supply and came back to Sewoon for testing. During the tests, LEE detected ripples around a capacitor connected to a 5V power source on the CPU board, something he had not noticed before. CHO asked him to replace it, as well as all the ICs on the board, even if they did not seem faulty.

Two days later, CHO and I went back to the church. Both KIM and PARK waited in the auditorium. We reinserted the three boards into the rack but left the backup power supply unconnected. We all moved to the organ's console. CHO switched it on, and it powered immediately. He tested if the memory, stop knobs, and keyboards worked fine. We tried again in a few minutes. It booted up immediately.

KIM: “Good god, what a relief!”

PARK: “We need a round of applause.”

CHO: “So it could’ve been the CPU board after all.”

PARK: “It’s fully repaired now, then?”

CHO: “Yes…but perhaps we should have the backup power supply connected just in case.”

ME: “You’ve been using the organ without any problem though, right?”

PARK: “Yeah, it’s been working, but we have weddings and a lot of other major events planned...we were worried that something would go wrong in the middle of an event because it hasn’t been fully fixed, we’d be in serious trouble.”

We all then moved back to the control room and installed the backup power supply for the organ. CHO told KIM and PARK that they could just plug in the backup power supply if the boot time became longer than 2-3 minutes. The repair was concluded, for now.

Analysis: repairedness as a contingent state. This vignette ties together several aspects that we have described so far regarding how a state of repairedness is typically contingent on many aspects other than an obvious yes-or-no judgment of whether the artifact under repair is working and the job is done. The conclusion of the repair job of the organ was contingent on CHO’s need to get paid, the pressures on the church to host impending events, and the impending change in board members, among other contingencies. A state of repairedness could not be objectively defined in advance and was subject to negotiations over properties and their sufficiency.

There was also serendipity — if it hadn't been for CHO's last inspection that led to the replacement of additional parts on the CPU board, the previously-defined state of repairedness would have been enough to conclude the repair. A "full" repair was initially aimed for and desired, but was not expected at this late stage by both the church and CHO. The organ was working sufficiently for the church, and both the church and CHO had a motivation to conclude the repair sooner than later.

Achieving a contingent state of repairedness meant the artifact had been repaired in the given repair situation, for the time being. In this vignette, while CHO said that the organ was now "fully repaired," CHO and the church agreed to install the backup power supply to be used when the boot time began to slow again. Their agreement acknowledged that the type and sufficiency of the properties relevant to the organ's repairedness could shift over time.

6.5 Discussion

The concept of repairedness contributes to CSCW research on repair and more broadly on the role of artifacts in collaborative work settings and the sustainability of artifacts. First, we summarize our findings and lay out our main contribution of the concept of repairedness as a way to frame the conclusion of repair work. Drawing on the lens of multiplicity from STS, we then discuss how our main contribution of this concept makes visible (i) the *multiple possible versions* of an artifact under repair and (ii) the process through which multiple versions of the artifact become *contingently stabilized* into a state of repairedness. We finally examine how repairedness may offer a distinct lens on issues of appropriation in research on the sustainability of artifacts.

6.5.1 Repairedness: when does an artifact under repair become a working artifact?

Much research in CSCW and STS has surfaced how the moments of breakdown in repair make visible the contingent nature of work practices and relations surrounding artifacts (e.g., Latour 2005; Suchman 2007; Star and Ruhleder 1996; Lutters and Ackerman 2002; Knorr-Cetina 2005; Mikalsen and Monteiro 2018). Our contribution to the CSCW literature on repair is to highlight and unpack the contingent, situated processes of negotiating the *conclusion* of repair work by introducing and framing a concept of "repairedness." In so doing, we extend the literature's emphasis on repair as contingent and ongoing accomplishments (Jackson 2017; Houston et al. 2016) by distinctly framing the contingencies involved in our setting in the *stabilization* of the multiple versions of artifacts under repair into working versions.

Our study unpacked how the set of properties of artifacts under repair at Sewoon were identified, prioritized, and evaluated for sufficiency for these artifacts to achieve repairedness. These properties were also scoped, re-prioritized, and temporarily substituted for throughout the repair process, shaping and re-shaping how a given artifact under repair was supposed to look and function for it to be considered working again. Challenges associated with negotiating repairedness also varied. In the case of the Braun record player, the properties to be considered for its repairedness were non-negotiable and straightforward, while concluding the repair of the electrical pipe organ involved a more complicated and drawn-out negotiation process.

While negotiating repairedness was also contingent on contextual factors such as the cost or complexity of repair work (e.g., Houston et al. 2016; Rosner and Ames 2014), our study highlighted how such negotiation also involves people's engagement with artifacts at the level of their properties in particular repair situations. Our vignettes showed how repairedness could be achieved by restoring *some* of the artifacts' original properties *to a degree* deemed sufficient to their owners. Repairedness, then, is important not just since the conclusion of repair has been less-examined, but since the process of negotiating this conclusion spotlights how the properties of artifacts themselves are contingent, an aspect of repair work that is often left in the background (Orlikowski and Iacono 2001).

6.5.2 Multiple versions of artifacts under repair and their contingent stabilization

In our vignettes, the process of achieving a state of repairedness highlighted the messy and contingent boundaries between artifacts under repair and working artifacts. We also noted possibilities of engaging with artifacts as multiple versions. The repairedness of the electric pipe organ, for instance, involved considering different versions of the artifact, such as one that booted up immediately and another that was free from the power and memory problems but whose boot time varied. A state of repairedness necessarily included certain properties of the artifact (e.g., being free of the static noise and having appropriate volume levels) while excluding others (e.g., supporting the bass sound) (Schubert 2019). Drawing on the lens of multiplicity in STS (e.g., Mol 2002; Law 2002), our concept of repairedness thus helps illustrate how people navigate through multiple versions of the same artifact by differently configuring its properties to negotiate and arrive at one version that renders it working again in a repair situation for the owner.

For repairedness to be achieved and a repair job to be concluded, the artifacts in our study needed to be stabilized as a set of properties at different levels of sufficiency — one version out of multiple possible versions. The stabilization, however, was neither fixed nor permanent, but itself contingent on the situation: both the owners and repair workers anticipated that the artifacts' properties and their contexts would change over time, requiring a different state of repairedness for any future repairs. The contingent stabilization of artifacts under repair, in a sense, points to the interplay between multiplicity and singularity in our engagement with artifacts, or what Law (2002, p.160, emphases original) calls the "double trick of managing the simultaneous performance of singularity and multiplicity, of *being* singular while *performing* multiplicity, or...of being multiple while performing singularity."

Viewing repairedness as the result of contingently stabilizing an artifact under repair into one version comprising a set of properties allowed us to provide descriptions at the micro-level of properties of not just how artifacts under repair are performed and enacted differently according to different coordinations of people, resources, and practices (e.g., Castellani et al. 2009; Houston 2019; Dye et al. 2019; Bjørn and Hertzum 2011), but of how such performances or enactments also draw on negotiating multiple versions of artifacts under repair themselves. By highlighting how concluding a repair job required negotiations about the artifact at the level of its properties, our study contributes to the analysis of the contingent role of properties of artifacts in repair work.

As in our findings, existing concepts of artifacts in CSCW — e.g., boundary objects, boundary negotiating artifacts, intermediary objects, coordinative artifacts, and knowledge artifacts (Star 2010; Lee 2007; Boujut and Blanco 2003; Schmidt and Wagner 2004; Cabitza et al. 2013; Bowers et al. 1995; Ackerman and Halverson 2004; Blomberg and Karasti 2013) — emphasize how artifacts have flexibility and are necessarily underspecified (Ackerman et al. 2013; Schmidt and Simone 1996; Suchman 2007). Artifacts are "plastic" but "robust", or "open" but "bounded" so that they can be recognized and shared as such while remaining adaptable (Bowker and Star 1999; Cabitza et al. 2013; Ackerman et al. 2013). Collaboratively maintaining an artifact that is both open and bounded, as Thomer et al. (2019, p.15) suggest, involves reflective and iterative processes of "enumerating" and "interrelating" its properties. Often viewing artifacts as tools or resources for helping people communicate and share representations within and across groups, however, extant research tends to focus on the contingent and multiple meanings that artifacts come to take on through diverse practices, while making implicit how the properties of artifacts play out in this process.

Consistent with a multiplicity perspective, our focus on the contingent stabilization of artifacts under repair points to the importance of ontological assumptions in theorizing artifacts in repair and other collaborative work practices. By shedding light on how the properties of an artifact are contingently stabilized in a state of repairedness, we draw attention to the need to also understand and support this process of contingent stabilization of the properties of the artifacts themselves. Next, we discuss why repairedness also is a useful and distinct lens for understanding the sustainability of artifacts.

6.5.3 Sustainability through repairedness

Jackson (2014, p.234) argues that a repair lens is critical to understanding how we generally work and live with artifacts in that it shifts our focus "from moments of production to moments of sustainability and the myriad forms of activity by which the shape, standing, and meaning of objects in the world is produced and sustained." He proposes a situated, contingent view of our sustainable interaction with artifacts grounded in "broken-world thinking," in which, repurposing Tolstoy's quote about unhappy families, "[a]ll broken technologies are broken in their own way" (p.228). Our study, by attending to when a repair job becomes concluded and an artifact under repair contingently stabilized, builds on his statement as follows: *just as there are many ways in which a working artifact becomes broken and in need of repair, there are also many ways in which an artifact under repair becomes repaired and working again*. By framing the concept of repairedness, we contribute a lens for broken-world thinking that identifies processes for sustaining interaction with artifacts by contingently stabilizing their properties into a working version.

As a lens on sustainability through broken-world thinking, repairedness sheds fresh light on the well-established CSCW concept of appropriation that is regarded as a key condition of sustainability (Meurer et al. 2018). Appropriation refers to the idea that the ways in which "technologies are adopted, adapted and incorporated into working practice" are often unexpected or unintended by their initial design (Dourish 2003, p.467; Tscheligi et al. 2014; Lindtner et al. 2012). Through appropriation, users of a technological artifact may use its existing properties differently across contexts, give the artifact a new meaning or function, or change practices associated with its use (Muller et al. 2016). As in our concept of repairedness, then, appropriation similarly concerns efforts towards making artifacts work as situations evolve (e.g., Balka and Wagner 2016; Orlikowski 1992; Pipek 1995; Draxler and Stevens 2011; Belin and Prié 2012).

If an appropriation lens focuses on how a designed artifact's meanings, properties, and associated practices are continually *changed* over time, a repairedness lens focuses on how the multiple versions of an artifact at any given time are negotiated into a *contingently stabilized* working version. The focus on changing the use of an initially designed artifact versus stabilizing multiple versions of an artifact points to how our repairedness lens may broaden CSCW's understandings of the contingent nature of appropriation processes.

First, several studies of sustainability and appropriation in CSCW have drawn on an evolutionary lens of technology design to make sense of the mechanisms of contingency (Jackson et al. 2014; Jackson 2017). Changes in artifacts are characterized as processes of *adaptation* that follow *paths* or *trajectories*, and whose effects play out over *lifecycles* (e.g., Belin and Prié 2012; Cohn 2016). A repairedness lens sheds light on more phenomenological aspects of sustainability — how the contingent properties of an artifact are negotiated on a moment-to-moment basis, and not just across more extended paths or trajectories. Stabilizing the artifact into one version (i) does not necessarily lead an artifact down a path or trajectory that implies closing off possibilities for other versions, and (ii) is a precarious state that will depend on a continued process of negotiation of the multiple versions that may be incoherent, divergent, or even contradictory (Mol 2002; Law and Singleton 2005).

Second, while studies of appropriation concern how artifacts may change in ways unexpected by their initial design, a repairedness lens focuses on how an artifact is made to work again in a way that appears to maintain its identity as a singular object. The properties considered for the artifacts at Sewoon were contingently surfaced and stabilized but for the most part were initially designed in. Thus, we adopt the *-ness* in repairedness to emphasize a phenomenological concern with how the properties of multiple versions of artifacts are stabilized as they shift back and forth from background to active concern, rather than how novel properties are evolved beyond the artifact's initial design. At Sewoon, for example, the *organ-ness* of the electric pipe organ was retained across multiple versions during the prolonged repair, drawing on different properties according to what made it a working organ in different situations. In this focus on stabilization across multiple versions of the artifact, repairedness is consistent with what Law (2002, p.2) calls "fractional coherence," or ways of "drawing things together without centering them."

Jackson (2014, p.232-3) asks: "What if we can build new and different forms of solidarity with our objects (and they with us)? And what if, beneath the nose of scholarship, this is what we do every day?" Seen through our lens of repairedness, one way towards designing for such sustainable solidarity with our artifacts may be to support the processes of negotiating the stabilization of artifacts into a version that works in a particular situation, for the time being. Our study showed how the repair workers and owners of artifacts at Sewoon engaged in these reflective and interactive processes of negotiation to enable the owners' relationships with their artifacts to continue. Making these negotiations were what the repair workers and owners did as part of their daily practices, and what our study hoped to capture and learn from.

6.5.4 Limitations and future work

We believe that our findings are likely to generalize to other settings, in different socio-historical contexts, in different work settings, and across many work practices. Nonetheless, there are some limitations to theorizing repairedness in the context of our site. Repair workers in this study were predominantly males in their 60s and 70s with decades of experience, who continued to work primarily with analog machines, and who operated their workshops independently. The site was situated in a particular work setting and socio-historical context in Seoul.

Future research would extend beyond our site to understand how processes of negotiating the repairedness of digital artifacts might be different from, or similar to, those of analog artifacts that we observed in this paper. What are the differences in attitudes to, or strategies for, negotiating repairedness within and across different cultures or demographics? And what are the implications of these differences for understanding processes of negotiating repairedness and, more broadly, for support for the sustainability of artifacts?

Given the often-prolonged and asynchronous timelines of repair work at Sewoon, this study is also limited in the number of artifacts for which the entire repair process could be observed during fieldwork. While the extended observation of chat rooms with CHO's customers allowed us to explore a wider range of repair projects, further research might consider relationships between artifact types (e.g., a CPU board in industrial machinery, specialized lab equipment, consumer electronics) or properties (e.g., aesthetic, functional, memory-related) and processes of negotiating repairedness.

Our analysis of how properties of an artifact under repair become contingently stabilized also points to possibilities for future work that theorizes and informs the conditions for such stabilization. Studies could examine how the properties of an artifact become iteratively identified, prioritized, and evaluated for sufficiency in a given repair situation. What are the relationships among multiple possible working versions (e.g., Mol 1999), given their different or overlapping properties? How could repair workers and owners of artifacts communicate different versions throughout the process of negotiating repairedness? These questions suggest that repairedness may be a rich lens for future research on repair work and artifacts as well as sustainable design.

6.6 Conclusion

Drawing on an ethnographic study of an analog electronics repair community in Seoul, South Korea, this chapter observed how the conclusion of repair work depended on negotiating a contingently stable working version out of multiple versions of an artifact under repair — what I refer to as a state of repairedness. To negotiate repairedness, repair workers and owners of artifacts identified, prioritized, and determined the sufficiency of particular properties of the artifacts in their repair situations in order to stabilize these properties into a version of a working artifact. This stabilization was contingent in the sense that shifts in the relevance of the artifacts' properties and contexts required the negotiation of a different state of repairedness. A repairedness lens points to opportunities for CSCW research to support sustainable engagement with artifacts by accounting for the properties underlying multiple versions of artifacts. Contingently stabilizing a set of properties to conclude repair work does not necessarily mean preserving artifacts' original functions or changing them into something else, but rather negotiating a version that works for their owners, for the time being.

In terms of an end-user approach to actionable data science systems, a repairedness lens sheds light on phenomenological aspects of data users' situated engagement with the properties of artifacts that are captured as data. Repairedness helps account for the contingency in the properties of the artifact itself (i.e., contingently identifying, evaluating, and negotiating the properties of the artifacts by situated repair workers to respond to a problematic situation). While the repair workers might not be able to be referred to as end-users of data science systems, they provided critical insights into the inherent multiplicity and uncertainty of artifacts that humans engage with, and challenges associated with selecting and putting together data on relevant properties to carve out one version of artifacts that are considered actionable and working.

7 Towards Actionable Data Science Systems: An End-user Approach

How can we make data science systems more actionable for their end-users? This dissertation has explored this question by developing an end-user approach, in which actionability depends on placing end-users and their data practices at the center of data science systems. My strategy has been not to study data science systems directly, but to study how end-users generally work with data as a basis for theorizing actionability in data science systems. Bringing together four studies on three different settings (craft brewers producing beers, people with visual impairments engaging with photos on their smartphones, and repair workers repairing broken artifacts), I have (i) framed participants from each of these settings as potential *end-users* of data science systems, and therefore as *data users*, (ii) unpacked participants' data practices, or the processes by which, as *data users*, they work with their data to engage with phenomena in their practices and settings, and (iii) explored how insights into end-users' data practices may inform the theory and design of data science systems that are actionable for end-users.

In this chapter, I first characterize end-users' data practices and summarize three themes across the four studies presented in chapters 3-6 for understanding data science systems in an end-user approach: (i) participants worked with their own data to take action; (ii) working with data depended on participants' ability to interact with conceptual and material constraints on artifacts and phenomena that they engaged with in their practices and settings; and (iii) accessing artifacts and phenomena was challenging. I then analyze how these findings contribute to understandings of actionability in data science systems. My contributions include (i) identifying the central role of properties of artifacts and phenomena to end-users' data practices; (ii) describing the processes by which end-users construct actionable versions of artifacts and things to take situated actions; (iii) explaining these processes theoretically through the idea of registering artifacts and phenomena through data; and (iv) characterizing the sustainability of these data practices as the ongoing contingent stabilization of artifacts and things. Using my analysis, I propose a set of requirements for an end-user approach to actionable data science systems of *relevance*, *flexibility*, and *accountability*.

7.1 Characterizing end-users' data practices

7.1.1 Participants used their own data to take action

Participants across the four studies used the data that they themselves collected or managed, and that were therefore situated in participants' own intentional practices and settings. This aspect of data use is different from data science research where data scientists or data science teams work with a dataset that is produced externally (e.g., provided by distant "data workers" or generated by users using a service). Humans who generate data for use by distant data scientists may not be aware of why they are collecting data or why their data are being otherwise collected; they may not work with their own data intentionally to understand or act on something. In such cases, it is difficult to consider the humans as potential end-users of data science systems. Emphasizing the importance of end-users using their own data is central to the perspective in this dissertation on the question of what it means for a data science system to be actionable. Actionability for whom and for what purposes? An online company that seeks to accurately predict patterns of its clients' behavior based on their use data?

For my participants, end users' data were a *resource for action* (e.g., Suchman 2007). They were aware that the data were about particular artifacts or phenomena of interest with which they wanted to engage in particular ways. The brewers recorded data on variables about a brew on their brewsheet to look for anomalies in their data and take actions to address them in their brewery. The people with visual impairments wanted to engage with the photos about particular events that they themselves took or received on their smartphones to jog their memory or reminisce with the photos. The repair workers measured the current of the circuit of a broken machine to use that data to decide what action to take on that machine in their repair workshops. The implication is that end-users need to be the owners of data for data science systems to be actionable in their practices and settings.

7.1.2 Participants took action by working with data about the properties of artifacts

The studies featured settings where participants took action by working with data about conceptual and material properties of artifacts and things in their phenomena of interest. Existing data science research tends to render the contingent properties of artifacts invisible or irrelevant by focusing on how data scientists work with mostly given datasets (e.g., by focusing on issues such as cognitive opacity of given models or the messiness of given datasets). For the brewers, using data depended on accounting for the biochemical and microbiological properties of ingredients and brews relevant to the phenomena of fermentation. For the repair workers, the physical properties of the parts and their connections in the artifact's original design posed formidable material constraints grounded in electrical engineering. For people with visual impairments, the data for generating image descriptions had to be put together in particular ways to construct a mental image that was appropriate, relevant, or meaningful to their intended photo activities.

Across diverse users and situations, participants in the four studies needed to consider pre-existing conceptual and material constraints on the artifacts and things that they interacted with in their practices and settings. A dataset in the abstract, without accounting for conceptual and material constraints specific to where the data were collected and how they were used risks producing insights that may not be actionable and thus inadequate for enabling end-users to become data users in data science systems.

7.1.3 Challenges of accessing artifacts and phenomena of interest

Participants in my studies commonly experienced challenges in gaining access to their artifacts and phenomena of interest through data. Challenges in accessing and making sense of artifacts and phenomena tend to be a marginal concern in technical data science research. Focusing on the analysis of datasets from external sources by data scientists leads to a view that a phenomenon may be understood through patterns and relationships *within* datasets, rather than through the relationship between data and phenomena from the perspective of humans using data situated in their phenomena of interest.

In the brewery, understanding a brew that entails a complex and dynamic microbiological process required more than simply recording data on a large number of variables regarding the brew. The brewers often had difficulty figuring out what a data point that deviated from an expected range meant in that particular brew, or connecting multiple variables to concretize the nature of the brew. In the repair study, repair workers needed to gain access to a broken machine, not just as a type of artifact with a predetermined list of properties and functions but as an idiosyncratic configuration of properties to be evaluated against a particular set of "sufficient" or "working" conditions (e.g., an electric pipe organ with a slow boot time (< 40 seconds) and with an unstable power source). The use of photos by people with visual impairments offered a more extreme case of participants having no or little direct access to the properties of artifacts or phenomena. The visual properties of photos needed to be translated into verbal, textual, or other non-visual data to be accessible to participants, creating confusion or frustration in their engagement with photos.

By highlighting the difficulty in accessing the artifacts and things in the phenomenon of interest through data, my findings suggest that an end-user approach to actionable data science systems depends on understanding the processes of using data to support such access.

7.2 Implications for understanding data science systems

7.2.1 The centrality of artifacts and phenomena in situating data, data users, and data science systems

Research on data science has tended to view data in terms of its statistical properties, or at best as informed by predefined ontologies of a domain (Ribes 2019). In its focus on large datasets, the data studied in such research are often data generated online. This focus on large, online datasets has popularized a view of data as "disembodied and place-less," and as something that can be understood on their own (Corple and Linabary 2020, p.155). In analyzing the role of data across my four studies, a theme that arose was how participants' actions depended on situating data in particular artifacts and phenomena of interest in their real-world practices and settings (e.g., Taylor et al. 2015; Lave 1984; Suchman 2007). When participants used data, the data were relevant to things that they could visualize, touch, or remember (e.g., the pH scale for this test batch, for a new seasonal beer, measured after the mash process; an image description stating "my mother smiling in front of the tree she planted").

Drawing on feminist materialism (e.g., Haraway 1988; Hayles 1999; Barad 2007), research in critical data studies problematizes the view of "data from nowhere" that has proliferated in data science research. They point out that, in a "data from nowhere" view, engaging with data in the abstract is tantamount to engaging with the world that the data supposedly capture. Working with a larger amount of data, it follows, leads to a richer and deeper understanding of this world (Corple and Linabary 2020; Bender et al. 2021; also Bowker 2013). By presenting settings where both data and the humans using the data “belon[g] somewhere” (Heidegger 1962, p.136), my dissertation highlights humans’ purposes for using data for engagement with phenomena through artifacts in the world. I studied how participants engaged with and acted on artifacts and phenomena as they existed or unfolded in real-world material practices and settings. The “belonging somewhere” of data and humans points to the importance of grounding data not only to obstacles in the physical environment (as discussed in robotics, e.g., e.g., Brooks 1990; Nilsson 2007), but to specific artifacts and phenomena in the physical environment (as discussed in research on phenomenology, e.g., Agre and Horswill 1997; Olsen 2010; Hartman 2011; Verbeek 2005).

The studies in this dissertation bring our attention to properties of specific artifacts and phenomena with which participants engaged. The physical or material properties of these situated artifacts and phenomena (e.g., visual elements and their compositions in photos, microbiological reactions in brewing) open up possibilities for relations situated in particular practices and settings, while closing off possibilities for others (Olsen 2010). Serres (1996) commented that human relationships are critically mediated by artifacts, and the relationships “would have been airy as clouds were there only contacts between subjects” (p.87). I argue that, without accounting for the effects of artifacts and phenomena, data and our relationships with data would be airy as clouds were there only contacts between data.

The situatedness of data and humans who use the data indicates the situatedness of data science systems. Data science systems are developed from data that reflect humans’ data practices, and will ultimately be deployed for the use of these humans in their practices and settings. In data science research, however, the role of these situated data users is often limited to informing the work of technical data scientists with their domain expertise and knowledge. My dissertation proposes bringing these *data users* to the center of imagining and developing data science systems. As data users, they are not simply a repository of experience and knowledge, but end-users of data science systems.

A growing body of research in data science emphasizes that data science systems need to be seen as sociotechnical, rather than purely technical systems. This research focuses on the situated and negotiated nature of the data practices of data scientists and data science teams, usually in corporate or scientific settings. (e.g., Passi and Sengers 2020). My dissertation extends this sociotechnical perspective on data science systems by making more central the properties of artifacts and phenomena relevant to end-users of these systems. In the end-user approach that I propose, data science systems depend on not only negotiations and practices regarding the meaning of data and the appropriateness of problems and analytical techniques. Data science systems also depend on engaging with the properties of artifacts and phenomena with which the data are concerned.

7.2.2 Registering actionable versions of artifacts and phenomena

Participants in my studies viewed their artifacts and phenomena of interest as inherently multiple or uncertain. Yet their ability to take action depended on constructing a working version relevant to a situation. The brewers drew on data on multiple variables about the brewing process in their paper brew sheets to get a sense of what was happening in a batch. They conceptualized a brew based on a particular subset of variables (e.g., the temperature and amount of water, the type or amount of malt used, pH values at different stages) that emerged as relevant to some aspect of the brew that they wanted to address (e.g., an out-of-range pH value at the end of runoff). The continual emergence of relevant properties of the brew led to multiple possible versions of the brew at any time. Yet the brewers needed to decide on one version that would allow them to take meaningful action (e.g., adding hot water at a particular point during the brewing process). The repair workers examined and conceptualized a broken machine in the context of multiple possible working versions of the machine. For their repair work to be completed, they needed to carve out and negotiate with the artifact's owner one working version that drew on a particular set of properties as well as their priorities and their levels of sufficiency (e.g., an amplifier with no static noise and proper volume levels but also with no bass support). The visually impaired photo users leveraged the multiplicity of photos by getting multiple descriptions of the same photo to enrich their photographic experience. Yet they also needed to settle on one version or mental image of a photo to perform a particular activity (e.g., using descriptions and other data to search for a particular photo taken with my father during our trip to France).

Participants' efforts to turn the *multiplicity* of a photo into one actionable *singularity* highlight that artifacts and phenomena cannot be simply "assumed" (Smith 2019, p.35), nor are they "atomic" without any "internal structure" (Smith 1998, p.230). Turning multiplicity into a singular version involved mechanisms of putting together and making sense of data on properties of artifacts and phenomena. Participants selected data that seemed relevant to or important for their intended activity or situation.

These processes of selecting relevant data can be characterized as processes of *registering* their artifacts and phenomena, or making the artifacts and phenomena "ontologically intelligible" in a way that seemed appropriate for their intentions and settings according to some conceptual and material coherence (Smith 2019, p.35-6; Smith 1998, p.282). The repair workers identified and prioritized properties of an artifact under repair relevant to a particular repair situation. They pieced together data on relevant properties to carve out a working version that made sense electronically and practically. The brewers drew on logical and causal relationships among variables to piece together a version of a specific brew that would allow them to make interventions. In photo use, the visually impaired participants went through a process of constructing mental images of their photos by putting together descriptions and other data. To make these mental images meaningful and useful for their intended photo activities, participants put together data of or about their photos in many ways (e.g., describing visual elements in a photo in one consistent direction, providing data at different levels of abstraction, or personalizing the data on photos).

7.2.3 Registering artifacts and phenomena through data

Processes of registering properties of artifacts and things through data use were geared toward obtaining ontological and epistemic access to these artifacts and phenomena that would be *sufficient* for taking action in a situation. In Dreyfus' (1992) terms, artifacts and phenomena "only need to be made as specific as the situation demands" (p.275). Participants in my settings were not aiming to get complete or precise explanations that could be applied to general situations (e.g., brewing principles in general; an objectively working machine; or a type of image description that would allow any visually impaired photo user to perform any intended activity). They were interested in engaging with specific artifacts and phenomena as they were presented and unfolded in particular situations. In data science research, this engagement corresponds to what is increasingly referred to as data science "in the wild", as opposed to data science in the "lab" (e.g., Beede et al. 2020; Sambasivan et al. 2021; Passi and Sengers 2020). The participants' registration processes were grounded in the data presented to, and used by, them in their settings, and were thus "local and contextual" (Pietsch 2022, p.39). Participants registered their artifacts and phenomena "only *approximately"* (*Ibid.*) in that there always were some properties or conditions that remained unaccounted for in a situation (Pietsch 2022). In addition to approaching artifacts and phenomena as “approximate entities” that had specific properties (McCarthy 2000), participants assumed and worked with their data themselves as approximations. They used their data as guidance, references, and clues with which they could make judgments about the status of artifacts. The precision of data or models in the statistical sense was not a central concern.

Registration processes concerned using data to get at an approximate, but materially and conceptually coherent understanding of artifacts and phenomena that would be sufficient for action. This focus on sufficiency in terms of registration processes differs from data science research that focuses on sufficiency in terms of the amount of data gathered or the level of predictive accuracy of models and algorithms. Smith (2019) argues that existing data science and AI systems tend to be “postregistrational” or “postontological” in their conventional use of “a prefigured set of registration schemes” for training these systems (p.139). My findings suggest that, for data science systems to be actionable “in the wild” (e.g., for end-users), they need to account for and support mechanisms for registering artifacts and phenomena through data. By registration mechanisms, I have in mind the very types of processes that end-users can perform on their artifacts by engaging with phenomena through data.

7.2.4 Contingent stabilization of properties of artifacts

In the above section, I explored how humans’ engagement with inherently multiple and uncertain artifacts and phenomena is achieved through a process of registering artifacts and phenomena through data to account for conceptual and material and constraints. Once an actionable version is registered, the artifact or phenomenon becomes “stabilized” in that particular situation (Smith 1998) - that is, it is treated as a working version for that particular instance. My studies highlight that this stabilization of the properties of an artifact is always contingent and tentative. Registration of artifacts and things as data by end-users involves contingently connecting a particular set of properties that are presented as relevant to them while leaving out other properties (e.g., Law and Singleton 2005). Artifacts and things “defy unique or even finite registration” (Smith 1998, p.290). As situations change, artifacts and things may call for different registration schemes, stabilizing into different versions (e.g., what constituted a working version of the electronic pipe organ changed through the drawn-out repair process; and the same photo came to have different meanings according to the intended photo activity drawing on its different aspects). Registering an artifact or phenomenon through data always faces “feisty boundaries, eruptive boundaries, boundaries that are challenged and negotiated and forever redrawn” (Smith 1998, p.370).

In my four studies, the properties of artifacts and phenomena were contingently stabilized into different versions as participants engaged with these artifacts and phenomena in different ways. Their engagements, however, retained the core identity of the artifacts (e.g., a brew, a photo, or an electric pipe organ). In this sense, the ongoing registration process was about working over time with multiple versions of an artifact, while keeping it largely the same (Jones and Yarrow 2013). As pointed out in chapter 6, this view of ongoingness is complementary to, but distinct from the well-established concept in CSCW of appropriation. Appropriation attends to how the initially designed artifacts’ meanings, properties, and associated practices evolve into something else in trajectories, typically over a longer-run time period. Registration, as I use it to analyze the findings from my settings, attends to how ongoing engagement with artifacts leverages their inherent multiplicity of properties to achieve a contingently coherent and actionable version at a particular moment and in a particular situation. My focus is on the moment-to-moment stabilization of the properties of artifacts — that is, on “contingent stabilization.” Contingent stabilization raises some interesting questions on how to support the collection and use of data about contingently stabilized artifacts and phenomena. A main point I have tried to make is that accounting for contingencies in artifacts' properties is difficult to do through the conventional data science approach of statistical analysis of data that assumes a mostly given set of properties and stability in relations between these properties.

Attending to the ongoing processes of contingently stabilizing the properties of artifacts highlights the need in data science activities to iterate not just between data scientists and domain experts, but between end-users and their artifacts and phenomena. This contribution introduces a phenomenological approach to data science research and more broadly to CSCW research on data science and data work. First, extant data science research generally focuses on iterations *within datasets* (by adding or removing instances, or modifying labels) or *between data and models* (e.g., Hohman et al. 2020). My studies suggest that accounting for the iterations between data and artifacts in developing models is important to enabling end-users to generate actionable insights with data science systems. Second, research on artifacts in CSCW focuses on their interpretive flexibility. The concept of interpretive flexibility, Rosenberger (2018) argues, tends to prioritize the *social practices and relationships* involving the artifacts over *ontological and phenomenological aspects* of the artifacts themselves. The settings in my studies showed how starting with data about artifacts' properties could mediate the kinds of relationships between humans and artifacts, or between data and artifacts, that are possible in the first place that could help with the actionability of data science systems for end-users.

7.3 Requirements for actionable data science systems

So far, I discussed the common findings across the four studies included in the dissertation and used these findings to frame an end-user approach to making data science systems more actionable. For data science systems to be more actionable, I have argued for an approach to theorizing and design that emphasizes how end-users use their data to engage with artifacts and things in their practices and settings. An end-user approach to actionable data science systems acknowledges how the relevance of artifacts and phenomena within real-world practices and settings is inherently uncertain and contingent on situations of engagement. End-users of data science systems use data to register and take actions regarding artifacts through processes of situating and putting together data to be conceptually and materially coherent in their practices and settings.

Based on my findings about end-users' data practices and their implications for understanding data science systems, l next propose three requirements for data science systems to be actionable. The actionability of a data science system depends on (i) the *relevance* of data for end-users (do data capture the relevant properties of artifacts and things of interest, and the relevant relations among these properties?); (ii) *flexibility* of data use for end-users (does the system enable end-users to flexibly use data to interact with artifacts and phenomena from different points of view and across different situations over time?); and (iii) the *accountability* of systems in the context of end-users' practices and settings (can the data science system properly account for how end-users experience and engage with real-world artifacts and phenomena as part of their practices and settings?).

7.3.1 Data capture relevant properties of artifacts and phenomena

Insight into a phenomenon in a world can be considered actionable if it has relevance to an agent's external world, or "external validity" (Argyris 1996, p.390). In the same vein, I propose, insights that are constructed in a data science system can be considered actionable when they are relevant to end-users' external practices and settings. Research in CSCW has begun to explore how the insights of data science systems can be made relevant to their practices, goals, values, and cultures (e.g., Sultana et al. 2021; Bratteteig and Verne 2018; Passi and Jackson 2017; Passi and Sengers 2020). Extending this research, my dissertation argues that the relevance of insights generated by a data science system depends on the relevance of data to the properties of artifacts and phenomena in end-users' practices and settings. Relevance concerns not just plausible statistical correlations of input data to output data, but the "phenomenological fit," or the fit between data and artifacts in the practices and settings in which data are used (Sambasivan et al. 2021). Does a dataset properly reflect how end-users think and work with the properties of artifacts and phenomena in their practices and settings? Do the insights generated in a data science system help end-users take meaningful and effective actions with respect to these properties?

Relevance understood as the phenomenological fit between data and artifacts is different from relevance understood as the statistical fit between data and models (e.g., Sambasivan et al. 2021). Data relevance in data science literature is often explored in terms of statistical feature selection, or "the process of obtaining a subset from an original feature set according to a certain feature selection criterion, which selects the relevant features of the dataset" (Cai et al. 2018, p.70). The relevance of a dimension of data is defined in terms of the impact of features (input variables) on model performance in algorithmic or statistical terms, such as cutting computation time or increasing predictive accuracy. Features are categorized as statistically "relevant, irrelevant, or redundant" (Khalid et al. 2014, p.372). Feature selection is viewed as a task for optimization, where the objective functions are to minimize the number of relevant features to be included in models and to maximize the statistical fit between data and models (Guyon and Elisseeff 2003; Blum and Langley 1997).

My studies in this dissertation suggest that focusing on the statistical relevance of data may lead to capturing artifacts and phenomena in terms of arbitrary properties (i.e., inconsistent with or irrelevant to the artifacts' internal structure of properties), or which seem irrelevant to end-users' real-world practices and settings. In chapter 4, for example, while the data scientist developed a highly predictive ML model on yield efficiency, the brewers had difficulty seeing how the predictions could help them take actions on the yield efficiency issue in their existing practices and settings. My dissertation implies that data science systems should leverage how end-users work with data about relevant properties of artifacts and things to make the insights generated by data science systems actionable for end-users.

7.3.2 End-users flexibly use data to register properties of artifacts and phenomena

A key requirement of actionability in any technological system is the ability to allow end-users to flexibly develop or modify their systems according to their needs over time (Paternò and Wulf 2017). Likewise, actionable data science systems need to allow end-users of data to flexibly use their data to make sense of and act on relevant properties of their artifacts and things as situations in their practices and settings unfold over time.

Accounting for change in existing data science research is also typified by the idea of statistical feature selection. Feature selection typically is discussed in this research as selecting a small set of statistically relevant features for given datasets and models. The datasets and models are implicitly assumed to capture relevant properties of artifacts and phenomena in the real world. Engaging with models built from the given datasets, in turn, are assumed to help users of data science systems engage with artifacts and phenomena in their real-world practices and settings. Properties of datasets and models that are relevant in statistical terms, however, need not have corresponding relevance to properties of artifacts and things as they are presented to end-users engaged in phenomena. We currently lack systematic methods to evaluate the relationship between datasets and models and the phenomena to which they are assumed to correspond (Freiesleben et al. 2022; Boge 2022). End-users are mostly limited to collecting or pre-processing given datasets and working with the outcomes of models (i.e., predictions), giving them little flexibility to explore the logical or material relationships among features. The ability to flexibly interact with models alone - or with data in the context of model performance - would not provide sufficient support for end-users to make sense of or act on their artifacts and phenomena. By using a largely pre-defined set of features and assuming that the initial model will hold, existing data science research tends not to account the contingent relevance of properties of artifacts and phenomena.

My studies highlighted how the data that participants collected as well as the properties of artifacts and phenomena that the participants saw as relevant were contingent on their situation. Smith (2019) stresses that any registration scheme is only appropriate in a particular situation, and the "intelligent" use of data depends on the ability to make: "a seasoned judgment…as to how the various data registered the world, how they can be soundly assessed, and what is required in order to integrate their different perspectives in a way that is accountable to the same underlying world" (p.90). A key concern for participants in making their data actionable was the ability to flexibly explore conceptual and material relations among data variables amid contingencies in their practices and settings. I propose that actionable data science systems need to allow end-users to explore multiple ways of registering properties of artifacts and things through data, beyond just making predictions.

7.3.3 Accountability of data science systems to end-users' practices and settings

For data science systems to be actionable in end-users' practices and settings, they need to go beyond just producing statistically accurate predictions of outcomes. Data science systems also need to be accountable (e.g., Passi and Jackson 2018). Accountability has become a much-discussed topic in many literatures dealing with data science and AI systems, and it is worth unpacking how it is defined and what the end-user approach that I am proposing implies for accountability.

Defined as "the answerability of actors for outcomes" in the contexts of social sciences and the law, accountability is a key requirement for any system producing outcomes with potential consequences for social justice (Kohli et al. 2018, p.3; Vecchione et al. 2021). Actors who operate such systems may be made to answer for their social consequences by measures such as liability assessment or enforcement (Kohli et al. 2018). In computer science, Kohli et al. (2018) observe that accountability has often been treated as "a kind of *trace property* of software systems" (p.4, emphasis original) - that is, tracing (e.g., through documentation) what inputs were used, how a program ran, what outcome the system produced and which actors in the system took what actions. This in turn tends to make the concept reliant on the designers' intentions of how their systems were supposed to work.

Accountability has recently emerged as a key research topic in data science and AI, usually in terms of conducting audits of algorithms. Consistent with the social, legal and computer science definitions of accountability above, algorithmic accountability refers to making the operators and designers of algorithms answerable for the potential social consequences of their use. The auditing literature is still emerging and there is much room for extension regarding the concept of accountability. Existing approaches in the auditing literature have been mostly focused on externally evaluating an algorithm's accountability with respect to predefined sets of requirements *post hoc*, or on managing risks based on industry standards or government regulations (Raji et al. 2020; Brown et al. 2021). Scholars also criticize a lack of consistency among or operationalizability of proposed audit frameworks, and the limited role that proposed frameworks assume for end-users and other stakeholders in the audit process (Mittelstadt 2019; Vecchione et al. 2021; Brown et al. 2021).

My dissertation relates accountability instead to a goal of making data science systems more actionable for end-users. Through my four studies, I found that end-users' main concern for their data use was to make sense of, and effectively act on, properties of artifacts and things that became relevant in their practices and settings. To frame my findings, I draw on phenomenological approaches to accountability that emphasize situated actions with systems in the real world (e.g., Smith 1998; Smith 2019; Agre and Horswill 1997; Suchman 2002; Suchman 2007). Smith (2019), for example, proposes a view of accountability that emphasizes the contexts in which humans are situated, rather than a predefined "world model" assumed in an algorithm:

> "*Context-sensitivity is not merely a question of having a world model*…No such model can be adequate to all potential circumstances. Rather, sensitivity to context requires being able to choose or develop a model (registration scheme) adequate to whatever circumstances are at hand, backed by resolute commitment to hold that model (scheme) accountable to the world it is a model of" (p.139-140, emphasis original).

Here, accountability concerns "hold[ing] the whole situation we are thinking about to account" (p.89), as situations unfold in the real world throughout our use of systems, and understanding the world as the result or achievement of contingent registrational activities. For him, accountability is centrally about "getting up out" of what is happening within computer programs or models and "hold[ing] [systems]things accountable to being in the world," by making them able to "genuinely refer to an object, assess ontological schemes, distinguish truth from falsity, respond appropriately to context, and shoulder responsibility" (p.xiii).

Consistent with Smith's view of accountability as driven by contexts and situations, my dissertation suggests that data science systems could be made more actionable by putting end-users at the center. End-users should be enabled to build or modify data science systems in ways that they themselves consider appropriate to how they experience and want to engage with artifacts and phenomena in their real-world practices and settings. I propose that, in an actionable data science system, accountability is part and parcel of how end-users imagine, use, and adapt the system over time as part of their practices and settings, more than a property of the system to be evaluated *post hoc* against an externally predefined set of standards and requirements. By focusing on how end-users experience and adapt a system as part of their ongoing, contingent practices and settings, I emphasize a view of accountability that is less about assessing liability than it is about empowering end-users to take action (Shneiderman 2020; Paternò and Wulf 2017; Fischer et al. 2004; Doherty and Doherty 2018). In the end-user approach that I propose, the accountability of a data science system is a result of making it more actionable.

8 Conclusion

As illustrated in the previous chapter, my studies dealt with settings where participants collected and used their own data themselves to take actions on artifacts and things as they engaged in phenomena of interest in their practices and settings. I framed my participants as themselves *data users* and thus, as potential *end-users* of data science systems. Recent data science research has focused on settings where end-users tend not to be the producers or collectors of data used in building data science systems, or where the artifacts or phenomena of interest are highly standardized or well-defined. Data are manipulated by technical data scientists as abstract representations and fed to mostly predefined models. These settings are found in highly-specialized or advanced research institutions and corporations that have the resources to set up an internal data science team or use external data science services, and which are typically concerned with making statistical predictions that average across large numbers of data points.

By examining data practices in less-standardized and more resource-constrained settings, my dissertation may open up opportunities to make data science systems more accessible and accountable across a wider range of settings. My proposed end-user approach to actionable data science may also provide useful insights for large institutions and corporations to increase the actionability of their data science systems by more richly integrating end-users (e.g., domain experts) with the work of their technical data scientists.

An important question for future research is then: How could actionable data science systems be designed based on an end-user approach? In this final chapter, I sketch out preliminary research directions, with a focus on how the end-user approach that I propose may be drawn on by data science researchers and practitioners. I suggest four types of support (flexible and efficient data annotations, exploring logical and material relationships in data, generating intermediate predictions, and developing data science strategies), as well as related research opportunities to make data science systems more actionable for end-users. I conclude with the limitations of this dissertation as well as broader implications of my end-user approach for future research beyond data science.

8.1 Future research opportunities: exploring the implications of an end-user approach for designing actionable data science systems

8.1.1 Accounting for contexts: supporting flexible and efficient data annotations

The end-user approach to data science systems that I have proposed in this dissertation emphasizes that datasets are collected and ultimately used in local contexts. Even data instances of the same value collected within the same setting could have different meanings to data users in different situations. The implication is that the end-users of data should be able to access and construct rich, fine-grained annotations or documentation of their data to capture the contexts of their use.

Supporting the annotation or documentation of contexts (e.g., the source, collection, use, and maintenance of datasets) has emerged as an important agenda to improve the actionability and accountability of data science and AI systems (Gebru et al. 2021; Bratteteig and Verne 2018; Bender et al. 2021; Muller and Strohmayer 2022). Scholars have proposed documentation of, for example, why a dataset was collected, how each data instance was acquired or measured, and what verification mechanisms were used. In existing research, the focus of documentation, however, has been mostly to increase the transparency or accountability of entire datasets, and has assumed that dataset creators are different from dataset users. In this research, the design implications are to support the documentation process for specialized creators of datasets. Gebru et al. (2021), for example, explain that the burden of creating detailed documentations for datasets means that careful consideration is needed of how documentation tasks could be incorporated into the existing work practices of dataset creators. To make the documentation process more efficient, Afzal et al. (2021) propose the use of automated reports on datasets including meta-data, data characteristics, quality assessment results and recommendations, data sources, and related policy restrictions.

I suggest that a data science system empower its end-users to create and use *finer-grained* annotations not about a dataset as a whole or at one time, but about small amounts of data, done by end-users *as they collect and use the data* in their practices and settings on an ongoing basis. As one possible support, designers may analyze existing annotations (their purposes and content types) that the data users in a particular setting make in their physical and digital data artifacts, and provide a list of annotation options at the level of data instances that the users can easily add while collecting or using data in a data science system. For example, data users could be able to tag "preset" to data points when they record preset values instead of actual values, color-code data points when they are viewed as presenting a particular type of problem, or mark data points that they think are worth revisiting for a reason that cannot be articulated at the time of collection or use. Future research may explore the relationships between the types of settings in which data are collected and used and the types of annotations that end-users make, as well as how accounting for such relationships in a data science system affects the actionability of the system for end-users.

8.1.2 Accounting for registration activities: supporting exploration of logical and material relationships in data subsets

Across the settings examined in this dissertation, a key concern for data users was how to use their data to make intelligible and take actions on inherently uncertain and contingent artifacts and phenomena of interest. The process of making artifacts and phenomena intelligible – i.e., registering (Smith 2019) – through data depended on data users' ability to draw on subsets of variables that they considered relevant to the particular situation. Making this data actionable arose from how the *data users* that I studied explored conceptual and material relationships among the variables to gain a sufficient understanding for the artifacts and phenomena to be actionable. The focus of data science research on exploring statistical patterns within given datasets provides little explicit support for exploring subsets of data variables according to user-defined logical or material relationships that change across situations.

An end-user approach to data science emphasizes the need for data science systems to help humans use data at a "human scale." Human scale means at a scale at which humans routinely think about, experience, and engage with things in their everyday lives (Long and Ye 2019; Beckwith and Sherry 2020). For a data science system to generate outcomes that are more actionable for end-users, it may, for example, be designed to enable end-users to explore relevant subsets of variables and evaluate their material and conceptual coherence (e.g., does this make sense in my setting or according to my theory?) based on constraints that they themselves define and refine contingent on the situations that they find themselves in. An actionable data science system would also allow end-users to build and modify simple models that reflect material and conceptual relationships that they have identified (e.g., Rudin 2019). Designs that empower end-users to construct models out of subsets of variables would also open up research opportunities to study both general and domain-specific characteristics in how end-users use data to register artifacts and phenomena across different settings. Such research could provide important design guidelines for designers of data science systems in particular domains.

8.1.3 Accounting for a course of action: generating intermediate predictions

Beyond documenting data and exploring relationships among data variables, another important design consideration for an end-user approach to data science systems is how to make predictions generated by these systems actionable for end-users. Predictions become less actionable when they lack relevance to end-users' practices and settings. Even relevant predictions become less actionable when end-users have difficulty determining how the predictions could be used in their practices and settings. Most existing predictive algorithms and models provide a single outcome given a set of input variables. The settings examined in my studies highlighted that end-users' situated data practices were tied to a *process* of working with artifacts and phenomena (e.g., in the brewery: transferring grains to the lauter tun, adding hot water to the grains, extracting wort, cooling, transferring the batch to a fermentation tank, monitoring the fermentation process, and storing the brew in a bright tank). The relevant problems (e.g., optimizing yield efficiency) involved multiple stages and many interdependent variables. A single outcome, however accurate, did not provide end-users with much practical or theoretical guidance over what actions needed to be taken or at which point of their processes of engaging with their artifacts (e.g., brews). Not having much to say about determining or modifying a course of action as things unfold, the value of the single prediction approach for end-users may be limited to *post-hoc* reflections.

The end-user approach proposed in this dissertation suggests that many of the predictions that a data science system is capable of producing only become actionable in the context of end-users' existing work processes of engaging with properties of artifacts and phenomena. One implication is that making predictions from data science systems more actionable may be supported by generating a series of predictions regarding intermediate goals, at different stages of end-users' work process, and which offer check-points for evaluating and modifying their course of action. Predicting intermediate goals may be done initially by end-users working with data scientists. Researchers could study how such collaboration would set new intermediate, multi-stage goals for which algorithms will be developed, and relevant variables that will need to be additionally collected in the end-users' practices and settings. Future research may explore how data science systems can support end-users to experiment with multi-stage predictions appropriate to their phenomena, and what kinds of challenges or opportunities this approach may create for end-users.

8.1.4 Accounting for changes over time: supporting the development of data science strategies

Most ML algorithms concern one-off applications on a given dataset (Polyzotis et al. 2017). Temporal aspects mostly concern how the algorithms can achieve "generalization" as the data for the application changes over time. For end-users, however, not only do their data practices change over time, so does how they will want to use a data science system. As they gain an experiential understanding of what the system can or cannot do, as well as how their data practices mediate the performance of the system, they will want to adapt the variables they collect and the problems they formulate.

The possible design and research directions suggested above (supporting flexible and efficient data annotations, the exploration of conceptual, material relationships in subsets of data, and the generation of intermediate predictions) may be used together for end-users to explore new ways of interacting with their data science systems over time. That is, data users will want to themselves construct longer-run data science *strategies*. For example, end-users could modify an existing model to account for changes in their data practices and thus the characteristics of their dataset. They may want to develop models and data practice strategies that they consider appropriate for new properties of an artifact or things in their phenomena of interest that they want to address. By shifting the focus of data science from models and algorithms to end-users' data practices, an end-user approach seeks to inform the design of support for a rich range of actions that end-users can take to modify data and data science systems over time. An end-user approach thus would benefit from longitudinal studies of data science systems across different settings to develop insight into opportunities and challenges for enabling end-users to construct strategies for using data science systems over time.

8.2 Limitations and future research beyond data science

This dissertation has several limitations. First, the studies in this dissertation are limited to ethnographic and qualitative methods to examine participants' data practices in their practices and settings. The settings of these studies were all in Korea, and the participants for each study were demographically idiosyncratic (e.g., the brewers in chapters 3 and 4 were all males in their 30s and 40s, and the repair workers in chapter 6 were mostly males in their 60s and 70s). The levels of participants' expertise and their domains of practice varied, ranging from highly-experienced experts in specialized areas (analog electronics repair workers) to practitioners with some practical, working knowledge that built on several years' of experience (craft brewers) and non-technical users in a more general domain (photo users with visual impairments). The cultural, social, and physical particularities of these settings and participants may limit the generalizability of the studies. When analyzed together, however, the multiple cases are intended to help in developing an initial theoretical framework that could be applied across diverse practices and settings (Eisenhardt 2021). Future research may continue to explore different types of data users and their data practices, which could be creatively combined and iteratively analyzed to develop and refine the end-user approach proposed in this dissertation.

This dissertation proposes an approach to data science systems based on the abductive theorizing of the four different studies. The focus of these studies was on how participants worked with their *data* to explore implications for theorizing and designing data science systems. Except for the case of chapter 4, my studies did not directly touch on developing data science systems for the participants, or on testing and evaluating data science tools in their settings. Exploring how the participants would actually engage with data science tools as their end-users, and how engaging with such tools would change over time, was beyond the scope of this dissertation. As mentioned earlier, this limitation points to future opportunities to use insights into end-users' data practices for actually building data science tools for specific end-users and evaluating the actionability of these tools.

More broadly, this dissertation concerns the sustainability of humans' data practices, or how humans continue to think, work, and live with data effectively and sustainably in "the dataverse" (Bowker 2013, p.167). Developing insights into how humans use a wide variety of data to register relevant properties of artifacts to engage in phenomena of interest may help us explore data practices as a multisensory experience (Pink 2011). Future research may explore how technologies including data science systems could provide support for multisensory engagement with data that goes beyond visual or textual support for data practices (e.g., dashboards and documentation). How do humans simultaneously experience, make sense of, and synthesize multiple types of data to enable richer relationships with the world? Overall, my dissertation suggests that we recenter the focus of data science research and development on how humans as potential end-users of data science systems may sustainably use these systems as they intended.

Bibliography

Mark S. Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human-Computer Interaction,* 15 (2-3), 179-203.

Mark S. Ackerman and Christine Halverson. 2004. Organizational memory as objects, processes, and trajectories: an examination of organizational memory in use. *Computer Supported Cooperative Work (CSCW),* 13 (2), 155-189.

Mark S. Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing knowledge and expertise: the CSCW view of knowledge management. *Computer Supported Cooperative Work (CSCW)* 22 (4-6), 531-573.

Dustin Adams, Sri Kurniawan, Cynthia Herrera, Veronica Kang, and Natalie Friedman. 2016. Blind photographers and VizSnap: a long-term study. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 201-208.

Shazia Afzal, C. Rajmohan, Manish Kesarwani, Sameep Mehta, and Hima Patel. 2021. Data readiness report. *IEEE International Conference on Smart Data Services,* 42-51.

Philip E. Agre. 1997. Towards critical technical practice: lessons learned in trying to reform AI. In Geoffrey C. Bowker, Susan Leigh Star, William Turner, and Les Gasser (Eds.), *Social science, technical systems, and cooperative work: bridging the great divide,* Erlbaum Associates, 131-158.

Philip E. Agre and Ian Horswill. 1997. Lifeworld analysis. *Journal of Artificial Intelligence Research*, 6, 111-145.

Syed Ishtiaque Ahmed, Steven J. Jackson, and Md Rashidujjaman Rifat. 2015. Learning to fix: knowledge, collaboration and mobile phone repair in Dhaka, Bangladesh. In *Proceedings of the 7th International Conference on Information and Communication Technologies and Development*, 1-10.

Cristina Alaimo and Jannis Kallinikos. 2021. Managing by data: algorithmic categories and organizing. *Organization Studies*, 42 (9), 1385-1407.

Francesca Alby and Cristina Zucchermaglio. 2009. Time, narratives and participation frameworks in software troubleshooting. *Computer Supported Cooperative Work (CSCW),* 18 (2-3), 129-146.

Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: the role of humans in interactive machine learning. *AI Magazine,* 35 (4), 105–120.

Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019a. Guidelines for human-AI interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 1-13.

Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019b. Software engineering for machine learning: a case study. *IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice*, 291-300.

David Andrzejewski, Xiaojin Zhu, and Mark Craven. 2009. Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In *Proceedings of the Annual International Conference on Machine Learning*, 25–32.

Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin. 2018. Certifiably optimal rule lists for categorical data. *Journal of Machine Learning Research*, 18 (234), 1-78.

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. 2015. VQA: visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, 2425-2433

Chris Argyris. Actionable knowledge: design causality in the service of consequential theory. *The Journal of Applied Behavioral Science*, 32 (4), 390-406.

Michela Arnaboldi. 2018. The missing variable in big data for social sciences: the decision-maker. *Sustainability* 10 (10), 3415.

M. Arnold, R. K. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilovic, R. Nair, K. N. Ramamurthy, A. Olteanu, D. Piorkowski, J. Tsay, and K. R. Varshney. 2019. Factsheets: increasing trust in AI services through supplier’s declarations of conformity. *IBM Journal of Research and Development*, 63 (4/5).

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion,* 58, 82-115.

Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. Developing a research agenda for human-centered data science. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 529-535.

Sinem Aslan, Sinem Emine Mete, Eda Okur, Ece Oktay, Nese Alyuz, Utku Ergin Genc, David Stanhill, and Asli Arslan Esme. 2017. Human expert labeling process (HELP): towards a reliable higher-order use stable labeling process and tool to assess student engagement. *Educational Technology*, 57 (1), 53–59.

Susan Athey. 2017. Beyond prediction: using big data for policy problems. *Science*, 355 (6324), 483-485.

Jan Auernhammer. 2020. Human-centered AI: the role of human-centered design research in the development of AI. *Synergy - DRS International Conference 2020.*

Jan Balata, Jan, Zdenek Mikovec, and Lukas Neoproud. 2015. Blindcamera: central and golden-ratio composition for blind photographers. In *Proceedings of the Multimedia, Interaction, Design and Innovation*, 1-8.

Ellen Balka and Ina Wagner. 2006. Making things work: dimensions of configurability as appropriation work. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work,* 229-238.

Charles Bamforth. 2003. *Beer: tap into the art and science of brewing*. Oxford and New York: Oxford University Press.

Nikola Banovic, Antti Oulasvirta, and Per Ola Kristensson. 2019. Computational modeling in human-computer interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1-7.

Karen Barad. 2007. *Meeting the universe halfway: quantum physics and the entanglement of matter and meaning*. Durham and London: Duke University Press.

Antoine Barczewski, Anastasia Bezerianos, and Nadia Boukhelifa. 2020. How domain experts structure their exploratory data analysis: towards a machine-learned storyline. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1-8.

Shaowen Bardzell, Daniela K. Rosner, and Jeffrey Bardzell. 2012. Crafting quality in design: integrity, creativity, and public sensibility. In *Proceedings of the Designing Interactive Systems*, 11-20.

Payam Barnaghi, Amit Sheth, and Cory Henson. 2013. From data to actionable knowledge: big data challenges in the web of things. *IEEE Intelligent Systems*, 28 (6), 6-11.

Roland Barthes. 1981. *Camera lucida: reflections on photography,* Richard Howard (trans.). New York, NY: Hill and Wang.

Eric P.S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, 68 (6), 1397-1410.

Howard S. Becker. 1978. Arts and crafts. *American Journal of Sociology,* 83 (4), 862-889.

Richard Beckwith and John Sherry. 2020. Scale and the gaze of a machine. *Ethnographic Praxis in Industry Conference Proceedings*, 2020 (1), 48-60.

Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-12.

Amaury Belin and Yannick Prié. 2012. DIAM: towards a model for describing appropriation processes through the evolution of digital artifacts. In *Proceedings of the Designing Interactive Systems Conference*, 645-654.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Yoshua Bengio, Yann Lecun, and Geoffrey Hinton. 2021. Deep learning for AI. *Communications of the ACM*, 64 (7), 58-65.

Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018a. Interdependence as a frame for assistive technology research and design. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 161-173.

Cynthia L. Bennett, Jane E, Martez E. Mott, Edward Cutrell, and Meredith Ringel Morris. 2018b. How teens with visual impairments take, edit, and share photos on social media. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-12.

Cynthia L. Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P. Bigham, Anhong Guo, and Alexandra To. 2021. "It's complicated": negotiating accessibility and (mis)representation in image descriptions of race, gender, and disability. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-19.

H. Russell Bernard and Gery W. Ryan. 2010. *Analyzing qualitative data: systematic approaches*. Thousand Oaks, CA: Sage.

Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 648-657.

Jeffrey P. Bigham, Samuel White, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Millerγ, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, 333-342.

Pernille Bjørn and Morten Hertzum. 2011. Artefactual multiplicity: a study of emergency-department whiteboards. *Computer Supported Cooperative Work (CSCW),* 20 (1-2), 93-121.

Jeanette Blomberg and Helena Karasti. 2013. Reflections on 25 years of ethnography in CSCW. *Computer Supported Cooperative Work (CSCW),* 22 (4-6), 373-423.

Avrim L. Blum and Pat Langley. 1997. Selection of relevant features and examples in machine learning. *Artificial intelligence,* 97 (1-2), 245-271.

Florian J. Boge. 2022. Two dimensions of opacity and the deep learning predicament. *Minds and Machines*, 32 (1), 43–75.

Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by data: nonprofits, social enterprises, and the consequences of data-driven work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 3608-3619.

Christine L. Borgman, Jillian C. Wallis, and Matthew S. Mayernik. 2012. Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work (CSCW),* 21 (6), 485-523.

Christine L. Borgman. 2016. *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA: MIT Press.

Albert Borgmann. 2015. Stability, instability, and phenomenology. In Robert Rosenberger and Peter-Paul Verbeek (Eds.), *Postphenomenological investigations: essays on human-technology relations.* Lanham, MD: Lexington Books, 247-252.

Claus Bossen, Kathleen H. Pine, Federico Cabitza, Gunnar Ellingsen, and Enrico Maria Piras. 2019. Data work in healthcare: an introduction. *Health Informatics Journal,* 25 (3), 465–474.

Jean-François Boujut and Eric Blanco. 2003. Intermediary objects as a means to foster co-operation in engineering design. *Computer Supported Cooperative Work (CSCW),* 12 (2), 205-219.

Nadia Boukhelifa, Marc-Emmanuel Perrin, Samuel Huron, and James Eagan. 2017. How data workers cope with uncertainty: a task characterisation study. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 3645-3656.

John Bowers, Graham Button, and Wes Sharrock. 1995. Workflow from within and without: technology and cooperative work on the print industry shop floor. In *Proceedings of the 4th European Conference on Computer-Supported Cooperative Work (ECSCW),* 51-66.

Geoffrey C. Bowker. 2005. *Memory practices in the sciences*. Cambridge, MA: MIT Press.

Geoffrey C. Bowker and Susan Leigh Star. 1999. *Sorting things out: classification and its consequences*. Cambridge, MA: MIT Press.

Geoffrey C. Bowker. 2013. Data flakes: an afterword to “raw data” is an oxymoron. In Lisa Gitelman (Ed.), *Raw data is an oxymoron.* Cambridge, MA: MIT Press, 167-171.

danah boyd and Kate Crawford. 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society,* 15 (5), 662-679.

Karen L. Boyd. 2021. Datasheets for datasets help ML engineers notice and understand ethical issues in training data. In *Proceedings of the ACM Human-Computer Interaction,* 5 (CSCW2), 438, 1-27.

Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2117–2126.

Erin Brady. 2015. Getting fast, free, and anonymous answers to questions asked by people with visual impairments. *SIGACCESS Accessibility and Computing*, 112, 16-25.

Tone Bratteteig and Guri Verne. "Does AI make PD obsolete? exploring challenges from artificial intelligence to participatory design." 2018. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2*, 1-5.

Eric Breck, Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. 2019. Data validation for machine learning. In *Proceedings of the Conference on Systems and Machine Learning*, 1-14.

Charles L. Briggs. 1986. *Learning how to ask: a sociolinguistic appraisal of the role of the interview in social science research.* Cambridge: Cambridge University Press.

Rodney A. Brooks. 1990. Elephants don’t play chess. *Robotics and Autonomous Systems*, 6, 3-15.

Bill Brown. 2001. Thing theory. *Critical Inquiry,* 28 (1), 1-22.

Shea Brown, Jovana Davidovic, and Ali Hasan. The algorithm audit: scoring the algorithms that score us. *Big Data & Society,* 8 (1), 1-8.

Jenna Burrell. 2016. "How the machine ‘thinks’: Understanding opacity in machine learning algorithms." *Big Data & Society,* 3 (1), 1-12.

Monika Büscher, Jacki O’Neill, and John Rooksby. 2009. Designing for diagnosing: introduction to the special issue on diagnostic work. *Computer Supported Cooperative Work (CSCW),* 18 (2-3), 109-128.

Federico Cabitza, Gianluca Colombo, and Carla Simone. 2013. Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities. *International Journal of Human-Computer Studies*, 71 (1), 24-45.

Federico Cabitza, Angela Locoro, Camilla Alderighi, Raffaele Rasoini, Domenico Compagnone, and Pedro Berjano. 2019. The elephant in the record: on the multiplicity of data recording work. *Health Informatics Journal,* 25 (3), 475-490.

Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. 2018. Feature selection in machine learning: a new perspective. *Neurocomputing,* 300, 70-79.

Longbing Cao. 2017. Data science: challenges and directions. *Communications of the ACM*, 60 (8), 59-68.

Daniel Carter and Dan Sholler. 2016. Data science on the ground: hype, criticism, and everyday work. *Journal of the Association for Information Science and Technology*, 67 (10), 2309-2319.

Stefania Castellani, Antonietta Grasso, Jacki O'Neill, and Frederic Roulland. 2009. Designing technology as an embedded resource for troubleshooting. *Computer Supported Cooperative Work (CSCW),* 18 (2-3), 199-227.

Zaira Cattaneo and Tomaso Vecchi. 2011. *Blind vision: the neuroscience of visual impairment*. Cambridge, MA: MIT Press.

Stevie Chancellor, Shion Guha, Jofish Kaye, Jen King, Niloufar Salehi, Sarita Schoenebeck, and Elizabeth Stowell. 2019. The relationships between data, power, and justice in CSCW research. In *Conference Companion Publication of the ACM Computer Supported Cooperative Work and Social Computing*, 102-105.

Akemi T. Chatfield, Vivian N. Shlemoon, Wilbur Redublado, and Faizur Rahman. 2014. Data scientists as game changers in big data environments. In *Proceedings of the Australasian Conference on Information Systems*, 1-11.

Angelos Chatzimparmpas, Rafael M. Martins, Ilir Jusufi, and Andreas Kerren. 2020. A survey of surveys on the use of visualization for interpreting machine learning models. *Information Visualization,* 19 (3), 207-233.

Amy Cheatle and Steven J. Jackson. 2015. Digital entanglements: craft, computation and collaboration in fine art furniture production. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 958-968.

John Cheney-Lippold. 2017. *We are data: algorithms and the making of our digital selves.* New York: New York University Press.

Joohee Choi and Yla Tausczik. 2017. Characteristics of collaboration in the emerging practice of open data analysis. In *Proceedings of the ACM conference on Computer Supported Cooperative Work,* 835-846.

Adele E. Clarke, Carrie Friese, and Rachel Washburn (Eds.). 2015. *Situational analysis in practice: mapping research with grounded theo*ry. Walnut Creek, CA: Left Coast Press.

Patricia Ticineto Clough, Karen Gregory, Benjamin Haber, and R. Joshua Scannell. 2015. The datalogical turn. In Phillip Vannini (Ed.), *Non-representational methodologies: re-envisioning research*. New York and London: Routledge, 146-164.

Dan Cosley, Victoria Schwanda Sosik, Johnathon Schultz, S. Tejaswi Peesapati, and Soyoung Lee. 2012. Experiences with designing tools for everyday reminiscing. *Human–Computer Interaction*, 27 (1–2), 75–198.

Marisa Leavitt Cohn. 2016. Convivial decay: entangled lifetimes in a geriatric infrastructure. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work,* 1511-1523.

Danielle J. Corple and Jasmine R. Linabary. 2020. From data points to people: feminist situated ethics in online big data research. *International Journal of Social Research Methodology,* 23 (2), 155-168.

Kristof Coussement and Dries F. Benoit. 2021. Interpretable data science for decision making. *Decision Support Systems,* 150, 113664.

Thomas H. Davenport and Rajeev Ronanki. 2018, Artificial intelligence for the real world. *Harvard Business Review*, 96 (1), 108-116.

Tijl De Bie, Luc De Raedt, José Hernández-Orallo, Holger H. Hoos, Padhraic Smyth, and Christopher KI Williams. 2022. Automating data science. *Communications of the ACM,* 65 (3), 76-87.

Marianne de Laet and Annemarie Mol. 2000. The Zimbabwe bush pump: mechanics of a fluid technology. *Social Studies of Science,* 30 (2), 225-263.

Jérôme Denis and David Pontille. 2015. Material ordering and the care of things. *Science, Technology & Human Values,* 40 (3), 338-367.

Jérôme Denis, Alessandro Mongili, and David Pontille. 2016. Maintenance and repair in science and technology studies. *TECNOSCIENZA: Italian Journal of Science & Technology Studies,* 6 (2), 5-16.

Vasant Dhar. 2011. Prediction in financial markets: the case for small disjuncts. *ACM Transactions on Intelligent Systems and Technologies,* 2 (3), 1-22.

Vasant Dhar. 2013. Data science and prediction. *Communications of the ACM,* 56 (12), 64-73.

Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. 2010. Mapping the landscape of sustainable HCI. In *Proceedings of the Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1975-1984.

Yvonne Dittrich, Dave W. Randall, and Janice Singer. 2009. Software engineering as cooperative work. *Computer Supported Cooperative Work (CSCW)*, 18 (5-6), 393-399.

Kevin Doherty and Gavin Doherty. 2018. Engagement in HCI: conception, theory and measurement. *ACM Computing Surveys*, 51 (5), 1-39.

Mateusz Dolata and Gerhard Schwabe. 2017. Paper practices in institutional talk: how financial advisors impress their clients. *Computer Supported Cooperative Work (CSCW)*, 26 (4), 769-805.

Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Paul Dourish, Annette Adler, Victoria Bellotti, and Austin Henderson. 1996. Your place or mine? Learning from long-term use of audio-video communication. *Computer Supported Cooperative Work (CSCW)*, 5 (1), 33-62.

Paul Dourish. 2003. The appropriation of interactive technologies: some lessons from placeless documents. *Computer Supported Cooperative Work (CSCW)*, 12 (4), 465-490.

Sebastian Draxler and Gunnar Stevens. 2011. Supporting the collaborative appropriation of an open software ecosystem. *Computer Supported Cooperative Work (CSCW)*, 20 (4-5), 403-448.

Hubert L. Dreyfus. 1992. *What computers still can't do: a critique of artificial reason*. Cambridge, MA: MIT Press.

Hubert L. Dreyfus. 2002. *A phenomenology of skill acquisition as the basis for a Merleau-Pontian nonrepresentational cognitive science*. https://philarchive.org/archive/DREAPO.

Hubert L. Dreyfus. 2007. Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, 171 (18), 1137–1160.

Ian Drosos, Titus Barik, Philip J. Guo, Robert DeLine, and Sumit Gulwani. 2020. Wrex: a unified programming-by-example interaction for synthesizing readable code for data scientists. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-12.

Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 297-307.

Johanna Drucker. 2018. Non-representational approaches to modeling interpretation in a graphical environment. *Digital Scholarship in the Humanities*, 33(2), 248-263.

Michaelanne Dye, David Nemer, Neha Kumar, and Amy S. Bruckman. 2019. If it rains, ask grandma to disconnect the nano: maintenance and care in Havana's StreetNet." In *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), 1-27.

James Edwards, Emory, Kyle Lewis Polster, Isabel Tuason, Emily Blank, Michael Gilbert, and Stacy Branham. 2021. "That's in the eye of the beholder": layers of interpretation in image descriptions for fictional representations of people with disabilities. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 1-14.

Kathleen M. Eisenhardt. 2021. What is the Eisenhardt method, really? *Strategic Organization*, 19 (1), 147-160.

Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, and others. 2015. From captions to visual concepts and back. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1473–1482.

Melanie Feinberg. 2017. A design perspective on data. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2952-2963.

Jacqueline Fennell and David M. Frohlich. 2005. *Beyond photographs: a design exploration of multi-sensory memorabilia for visually impaired people*. HP Labs Technical Report No. HPL-2005-151.

Juliana Jansen Ferreira and Mateus de Souza Monteiro. 2020. Do ML experts discuss explainability for AI systems? A discussion case in the industry for a domain-specific solution. *arXiv preprint arXiv:2002.12450*. Accessed 20 January 2022.

Gerhard Fischer, Elisa Giaccardi, Yunwen Ye, Alistair G. Sutcliffe, and Nikolay Mehandjiev. 2004. Meta-design: a manifesto for end-user development. *Communications of the ACM*, 47 (9), 33-37.

Gerhard Fischer, Kumiyo Nakakoji, and Yunwen Ye. 2009. Metadesign: guidelines for supporting domain experts in software development. *IEEE Software*, 26 (5), 37-44.

Oliver J. Fisher, Nicholas J. Watson, Josep E. Escrig, Rob Witt, Laura Porcu, Darren Bacon, Martin Rigley, and Rachel L. Gomes. 2020. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. *Computers & Chemical Engineering*, 140, 106881, 1-14.

George Fletcher, Paul Groth, and Juan Sequeda. 2020. Knowledge scientists: unlocking the data-driven organization. *arXiv preprint arXiv:2004.07917*. Accessed 1 January 2022.

Luciano Floridi. 2011. *The philosophy of information.* Oxford and New York: Oxford University Press.

Malcolm R. Forster. 2002. Predictive accuracy as an achievable goal of science. *Philosophy of Science* 69 (S3), S124-S134.

Timo Freiesleben, Gunnar König, Christoph Molnar, and Alvaro Tejero-Cantero. 2022. Scientific inference with interpretable machine learning: analyzing models to learn about real-world phenomena." *arXiv preprint arXiv:2206.05487*.

William R. Frey, Desmond U. Patton, Michael B. Gaskell, and Kyle A. McGregor. 2020. Artificial intelligence and inclusion: formerly gang-involved youth as domain experts for analyzing unstructured Twitter data. *Social Science Computer Review,* 38 (1), 42-56.

Batya Friedman and David G. Hendry. 2019. *Value sensitive design: shaping technology with moral imagination*. Cambridge, MA: MIT Press.

David Frohlich, Paul Drew, and Andrew Monk. 1994. Management of repair in human-computer interaction. *Human–Computer Interaction,* 9 (3-4), 385-425.

David M. Frohlich, Steven Wall, and Graham Kiddle. 2013. Rediscovery of forgotten images in domestic photo collections. *Personal and Ubiquitous Computing,* 17 (4), 729-740.

Joao Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. 2004. Learning with drift detection. *Brazilian Symposium on Artificial Intelligence*, 286–295.

Chuang Gan, Zhe Gan, Xiaodong He, Jianfeng Gao, and Li Deng. 2017. Stylenet: generating attractive visual captions with styles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 3137–3146.

Clément Gautrais, Yann Dauxais, Stefano Teso, Samuel Kolb, Gust Verbruggen, and Luc De Raedt. 2021. Human-machine collaboration for democratizing data science. In Stephen Muggleton and Nick Charter (Eds.), *Human-like machine intelligence*, Oxford: Oxford University Press, 379-402.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Communications of the ACM*, 64 (12), 86-92.

Yolanda Gil. 2017. Thoughtful artificial intelligence: forging a new partnership for data science and scientific discovery. *Data Science,* 1 (1-2), 119-129.

Yolanda Gil, James Honaker, Shikhar Gupta, Yibo Ma, Vito D'Orazio, Daniel Garijo, Shruti Gadewar, Qifan Yang, and Neda Jahanshad. 2019. Towards human-guided machine learning. In *Proceedings of the International Conference on Intelligent User Interfaces*, 614-624.

Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining explanations: an overview of interpretability of machine learning. In *IEEE International Conference on Data Science and Advanced Analytics,* 80-89.

Lisa Gitelman (Ed.). 2013. *Raw data is an oxymoron*. Cambridge, MA: MIT Press.

Lisa Gitelman and Virginia Jackson. 2013. Introduction. In Lisa Gitelman (Ed.), *Raw data is an oxymoron*. Cambridge, MA: MIT Press, 1-14.

Barney G. Glaser and Anselm L. Strauss. 2009. *Discovery of grounded theory: strategies for qualitative research*. Transaction Publishers.

Connie Golsteijn, Elise van den Hoven, David Frohlich, and Abigail Sellen. 2014. Reflections on craft research for and through design. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction*, 421-430.

Claudia Gonzalez Viejo, Damir D. Torrico, Frank R. Dunshea, and Sigfredo Fuentes. 2019. Development of artificial neural network models to assess beer acceptability based on sensory properties using a robotic pourer: a comparative model approach to achieve an artificial intelligence system. *Beverages*, 5 (2), 33, 1-10.

Stephen Graham and Nigel Thrift. 2007. Out of order: understanding repair and maintenance. *Theory, Culture & Society,* 24 (3), 1-25.

Thomas R.G. Green and Marian Petre. 1996. Usability analysis of visual programming environments: a 'cognitive dimensions' framework. *Journal of Visual Languages & Computing*, 7 (2), 131-174.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM Computing Survey*, 51 (5), 93, 1-42.

Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption crawler: enabling reusable alternative text descriptions using reverse image search. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 518, 1-11.

Danna Gurari, Yinan Zhao, Meng Zhang, and Nilavra Bhattacharya. 2020. Captioning images taken by people who are blind. *European Conference on Computer Vision*, 417-434.

Isabelle Guyon and André Elisseeff. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research,* 3, 1157-1182.

Alon Halevy, Peter Norvig, and Fernando Pereira. 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems,* 24 (2), 8-12.

Margot Hanley, Solon Barocas, Karen Levy, Shiri Azenkot, and Helen Nissenbaum. 2021. Computer vision and conflicting values: describing people with automated alt text. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 543-554.

Susumu Harada, Daisuke Sato, Dustin W. Adams, Sri Kurniawan, Hironobu Takagi, and Chieko Asakawa. 2013. Accessible photo album: enhancing the photo sharing experience for people with visual impairment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2127-2136.

Donna Haraway. 1991. *Simians, cyborgs and women: the reinvention of nature.* New York: Routledge.

Graham Harman. 2011. *The quadruple object*. Winchester, UK: Zero Books.

Douglas Harper. 1987. *Working knowledge: skill and community in a small shop*. Chicago: University of Chicago Press.

N. Katherine Hayles. 1999. How we became posthuman: virtual bodies in cybernetics, literature, and informatics. Chicago: University of Chicago Press.

Xiaodong He and Li Deng. 2017. Deep learning for image-to-text generation: a technical overview. *IEEE Signal Processing Magazine*, 34 (6), 109-116.

Christian Heath and Paul Luff. 1996. Documents and professional practice: "bad" organisational reasons for "good" clinical records. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 354-363.

Jutta Heckhausen and Richard Schulz. 1995. A life-span theory of control. *Psychological Review*, 102 (2), 284-304.

Richard Heersmink. 2021. Varieties of artifacts: embodied, perceptual, cognitive, and affective. *Topics in Cognitive Science*, 13 (4), 573-596.

Martin Heidegger. 1962. *Being and time*. New York: Harper and Row.

Martin Heidegger. 1971. *Poetry, language, thought*. New York: Harper and Row.

Christopher R. Henke. 1999. The mechanics of workplace order: toward a sociology of repair. *Berkeley Journal of Sociology*, 55-81.

Christopher R. Henke. 2017. The sustainable university: repair as maintenance and transformation. *Continent*, 6 (1), 40-45.

T. Ryan Hoens, Robi Polikar, and Nitesh V. Chawla. 2012. Learning from streaming data with concept drift and imbalance: an overview. *Progress in Artificial Intelligence*, 1 (1), 89-101.

Fred Hohman, Kanit Wongsuphasawat, Mary Beth Kery, and Kayur Patel. 2020. Understanding and visualizing data iteration in machine learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-13.

Md. Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga. 2019. A comprehensive survey of deep learning for image captioning. *ACM Computing Survey,* 51 (6), 1-36.

Youyang Hou and Dakuo Wang. 2017. Hacking with NPOs: collaborative analytics and broker roles in civic data hackathons. In *Proceedings of the ACM on Human-Computer Interaction,* 1 (CSCW), 1-16.

James S. Hough. 1991. *The biotechnology of malting and brewing*. Cambridge: Cambridge University Press.

Lara Houston and Steven J. Jackson. 2016. Caring for the" next billion" mobile handsets: opening proprietary closures through the work of repair. In *Proceedings of the 8th International Conference on Information and Communication Technologies and Development*, 1-11

Lara Houston, Steven J. Jackson, Daniela K. Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. 2016. Values in repair. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1403-1414.

Lara Houston. 2019. Mobile phone repair knowledge in downtown Kampala: local and trans-local circulations. Ignaz Strebel, Alain Bovet, and Philippe Sormani (Eds.). *Repair work ethnographies*. Singapore: Palgrave Macmillan, 129-160.

James Howison and James D. Herbsleb. 2011. Scientific software production: incentives and collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 513-522.

Paul Hughes. 2009. Beer flavor. In Charles W. Bamforth (Ed.). *Beer: a qualitative perspective*. Burlington, MA: Academic Press, 61-83.

Jina Huh, Kevin Nam, and Nikhil Sharma. 2010. Finding the lost treasure: understanding reuse of used computing devices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1875-1878.

Edmund Husserl. 2006. *Phantasy, image consciousness, and memory (1898-1925), vol. 11*. Springer Science & Business Media.

Nick Hynes, D. Sculley, and Michael Terry. 2017. The data linter: lightweight, automated sanity checking for ml data sets. *Machine Learning Systems Workshop at the Conference on Neural Information Processing Systems,* 1-7.

Marco Iansiti and Karim R. Lakhani. 2020. *Competing in the age of AI: strategy and leadership when algorithms and networks run the world*. Boston, MA: Harvard Business Review Press.

Masakazu Iwamura, Naoki Hirabayashi, Zheng Cheng, Kazunori Minatani, and Koichi Kise. 2020. VisPhoto: photography for people with visual impairment as post-production of omni-directional camera image. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1-9

Steven J. Jackson, Alex Pompe, and Gabriel Krieshok. 2012. Repair worlds: maintenance, repair, and ICT for development in rural Namibia. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 107-116.

Steven J. Jackson. 2014. Rethinking repair. In Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (Eds.). *Media technologies: essays on communication, materiality and society*. Cambridge, MA: MIT Press, 221-239.

Steven J. Jackson and Laewoo Kang. 2014. Breakdown, obsolescence and reuse: HCI and the art of repair. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 449-458.

Steven J. Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The policy knot: re-integrating policy, practice and design in CSCW studies of social computing. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 588-602.

Steven J. Jackson. 2017. Speed, time, infrastructure: temporalities of breakdown, maintenance, and repair. Judy Wajcman and Nigel Dodd (Eds.). *The sociology of speed: digital, organizational, and social temporalities*. Oxford: Oxford University Press, 169-185.

Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 203-210.

Yushi Jing, David Liu, Dmitry Kislyuk, Andrew Zhai, Jiajing Xu, Jeff Donahue, and Sarah Tavel. 2015. Visual search at Pinterest. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1889-1898.

Marina Jirotka, Charlotte P. Lee, and Gary M. Olson. 2013. Supporting scientific collaboration: methods, tools and concepts. *Computer Supported Cooperative Work (CSCW)*, 22 (4-6), 667-715.

Siân Jones and Thomas Yarrow. 2013. Crafting authenticity: an ethnography of conservation practice. *Journal of Material Culture*, 18 (1), 3-26.

Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2021a. Communicating visualizations without visuals: investigation of visualization alternative text for people with visual impairments. *IEEE Transactions on Visualization and Computer Graphics*, 28 (1), 1095-1105.

Ju Yeon Jung, Tom Steinberger, John L. King, and Mark S. Ackerman. 2021b. Negotiating repairedness: how artifacts under repair become contingently stabilized. In *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2), 328, 1-29.

Ju Yeon Jung, Tom Steinberger, John L. King, and Mark S. Ackerman. 2022a. How domain experts work with data: situating data science in the practices and settings of craftwork. In *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW1), 58, 1-29.

Ju Yeon Jung, Tom Steinberger, Junbeom Kim, and Mark S. Ackerman. 2022b. “So what? What’s that to do with me?” Expectations of people with visual impairments for image descriptions in their personal photo activities. *Designing Interactive Systems Conference*, 1-14.

Ju Yeon Jung, Tom Steinberger, and Chaehan So. 2022c. [Manuscript under submission]. Domain experts as owners of data: towards sustainable data science.

Hernisa Kacorri, Kris M. Kitani, Jeffrey P. Bigham, and Chieko Asakawa. 2017. People with visual impairment training personal object recognizers: feasibility and challenges. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 5839-5849.

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer. 2012. Enterprise data analysis and visualization: an interview study. *IEEE Transactions on Visualization and Computer Graphics*, 18 (12), 2917-2926.

Helena Karasti, Karen S. Baker, and Eija Halkola. 2006. Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long-term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW)*, 15 (4), 321-358.

Helena Karasti, Karen S. Baker, and Florence Millerand. 2010. Infrastructure time: long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW),* 19 (3), 377-415.

Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of Fairness, Accountability, and Transparency,* 1-10.

Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. 2017. Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering,* 29 (10), 2318-2331.

Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.

Charles Keller and Janet D. Keller. 1993. Thinking and acting with iron. In Seth Chaiklin and Jean Lave (Eds.). *Understanding practice: perspectives on activity and context*. Cambridge: Cambridge University Press, 125-143.

Mary Beth Kery, Bonnie E. John, Patrick O'Flaherty, Amber Horvath, and Brad A. Myers. 2019. Towards effective foraging by data scientists to find past analysis choices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-13.

Samina Khalid, Tehmina Khalil, and Shamila Nasreen. 2014. A survey of feature selection and feature extraction techniques in machine learning. *IEEE Science and Information Conference*, 372-378.

Gyeong-eun Kim. 2020. The IPO of Jeju Beer...The domestic craft beer market outlook. Maeil Business Newspaper. Available: https://moneys.mt.co.kr/news/mwView.php?no=2020121114448068387 (in Korean). Accessed 1 February 2021.

Hyun K. Kim, Sung H. Han, Jaehyun Park, and Joohwan Park. 2017. The interaction experiences of visually impaired people with assistive technology: a case study of smartphones. *International Journal of Industrial Ergonomics*, 55, 22-33.

Tim Kindberg, Mirjana Spasojevic, Rowanne Fleck, and Abigail Sellen. 2005. The ubiquitous camera: an in-depth study of camera phone use. *IEEE Pervasive Computing,* 4 (2), 42-50.

Rob Kitchin. 2014a. Big data, new epistemologies and paradigm shifts. *Big Data & Society,* 1 (1), 1-12.

Rob Kitchin. 2014b. *The data revolution: big data, open data, data infrastructures, and their consequences*. LA and London: Sage.

Tiffany Knearem, Xiying Wang, Jiaan Wan, and John M. Carroll. 2019a. Crafting in a community of practice: resource sharing as key in supporting creativity. In *Proceedings of the 2019 on Creativity and Cognition*, 83-94.

Tiffany Knearem, Xiying Wang, and John M. Carroll. 2019b. How professionals support amateurs' creativity within the brewing community. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work*, 252-256.

Karin Knorr-Cetina. 2005. Objectual practice. Karin Knorr-Cetina, Theodore R. Schatzki, and Eike Von Savigny (Eds.). *The practice turn in contemporary theory*. London and New York: Routledge, 184-197.

Martin Koehler, Alex Bogatu, Cristina Civili, Nikolaos Konstantinou, Edward Abel, Alvaro A.A. Fernandes, John Keane, Leonid Libkin, and Norman W. Paton. 2017. Data context informed data wrangling. *IEEE International Conference on Big Data*, 956-963.

Laura Koesten, Kathleen Gregory, Paul Groth, and Elena Simperl. 2021. Talking datasets–understanding data sensemaking behaviours. *International Journal of Human-Computer Studies*, 146, 102562.

Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. 2020. Mapping out human-centered data science: methods, approaches, and best practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work*, 151-156.

Nitin Kohli, Renata Barreto, and Joshua A Kroll. 2018. Translation tutorial: a shared lexicon for research and practice in human-centered software systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 1-7.

Stephen M. Kosslyn, William L. Thompson, and Giorgio Ganis. 2006. *The case for mental imagery*. Oxford: Oxford University Press.

Sanjay Krishnan, Daniel Haas, Michael J Franklin, and Eugene Wu. 2016. Towards reliable interactive data cleaning: a user survey and recommendations. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, 1–5.

Herbert M. Kritzer. 2007. Toward a theorization of craft. *Social & Legal Studies,* 16 (3), 321-340.

Josua Krause, Adam Perer, and Kenney Ng. 2016. Interacting with predictions: visual inspection of black-box machine learning models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 5686-5697.

Martin Kurze. 1996. TDraw: a computer-based tactile drawing tool for blind people. In *Proceedings of the ACM Conference on Assistive Technologies*, 131-138.

Stacy Kuznetsov, Christina J. Santana, and Elenore Long. 2016. Everyday food science as a design space for community literacy and habitual sustainable practice. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1786-1797.

Bruno Latour. 2005. *Reassembling the social.* Oxford: Oxford University Press.

Jean Lave. 1984. The values of quantification. *The Sociological Review*, 32, 88-111.

John Law. 2002. *Aircraft stories: decentering the object in technoscience*. Durham and London: Duke University Press.

John Law and Vicky Singleton. 2005. Object lessons. *Organization*, 12 (3), 331-355.

Charlotte P. Lee. 2007. Boundary negotiating artifacts: unbinding the routine of boundary objects and embracing chaos in collaborative work. *Computer Supported Cooperative Work (CSCW),* 16 (3), 307-339.

Sooyeon Lee, Madison Reddie, Chun-Hua Tsai, Jordan Beck, Mary Beth Rosson, and John M. Carroll. 2020. The emerging professional practice of remote sighted assistance for people with visual impairments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-12.

Sabina Leonelli. 2015. What counts as scientific data? A relational framework. *Philosophy of Science*, 82 (5), 810–821.

Sabina Leonelli. 2016. The philosophy of data. In Luciano Floridi (Ed.), *The Routledge handbook of philosophy of information*. London and New York: Routledge, 191-202.

Tuck Wah Leong, Richard Harper, and Tim Regan. 2011. Nudging towards serendipity: a case with personal digital photos. In *Proceedings of the BCS Conference on Human Computer Interaction*, 25, 385-394.

Bruno Lepri, Nuria Oliver, Emmanuel Letouze, Alex Pentland, and Patrick Vinck. 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31 (4), 611–627.

Michael J. Lewis and Tom W. Young. 2001. *Brewing*. New York, NY: Springer Science & Business Media.

Q. Vera Liao and Michael Muller. 2019. Enabling value-sensitive AI systems through participatory design fictions. *arXiv preprint arXiv:1912.07381*.

Jongho Lim, Yongjae Yoo, Hanseul Cho, and Seungmoon Choi. 2019. Touchphoto: enabling independent picture taking and understanding for visually-impaired users. *International Conference on Multimodal Interaction*, 124-134.

Luis Lima, Tiago Brandão, Nelson Lima, and José António Teixeira. 2011. Comparing the impact of environmental factors during very high gravity brewing fermentations. *Journal of the Institute of Brewing*, 117 (3), 359-367.

Siân E. Lindley. 2012. Before I forget: from personal memory to family history. *Human–Computer Interaction*, 27 (1–2), 13–36.

Silvia Lindtner, Ken Anderson, and Paul Dourish. 2012. Cultural appropriation: information technologies as sites of transnational imagination. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 77-86.

Silvia Lindtner. 2014. Hackerspaces and the Internet of Things in China: how makers are reinventing industrial production, innovation, and the self. *China Information*, 28 (2), 145-167.

Jessica Lingel and Tim Regan. 2014. "It's in your spinal cord, it's in your fingertips": practices of tools and craft in building software. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 295-304.

Jiali Liu, Nadia Boukhelifa, and James R. Eagan. 2020. Understanding the role of alternatives in data analysis practices. *IEEE Transactions on Visualization and Computer Graphics*, 2 (1), 66-76.

Steve Lohr. 2014. For big-data scientists, 'janitor work' is key hurdle to insights. *The New York Times.* https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html. Accessed 15 December 2021.

Ying Long and Yu Ye. 2019. Measuring human-scale urban form and its performance. *Landscape and Urban Planning,* 191, 103612.

Raoni Lourenço, Juliana Freire, and Dennis Shasha. 2019. Debugging machine learning pipelines. In *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning,* 1–10.

Thomas Ludwig, Volkmar Pipek, and Peter Tolmie. 2018. Designing for collaborative infrastructuring: supporting resonance activities. In *Proceedings of the ACM on Human-Computer Interaction,* 2 (CSCW), 113, 1-29.

Paul Luff, Jon Hindmarsh, and Christian Heath. 2000. *Workplace studies: recovering work practice and informing system design*. Cambridge: Cambridge University Press.

Jasmina Lukinac, Kristina Mastanjević, Krešimir Mastanjević, Gjore Nakov, and Marko Jukić. 2019. Computer vision method in beer quality evaluation—a review. *Beverages,* 5 (2), 38.

Wayne G. Lutters and Mark S. Ackerman. 2002. Achieving safety: a field study of boundary objects in aircraft technical support. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 266-275.

Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How data scientists work together with domain experts in scientific collaborations: to find the right answer or to ask the right question? In *Proceedings of the ACM on Human-Computer Interaction,* 3 (GROUP), 237, 1-23.

Haley MacLeod, Cynthia L Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding blind people's experiences with computer-generated captions of social media images. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 5988–5999.

Leah Maestri and Ron Wakkary. 2011. Understanding repair as a creative process of everyday design. In *Proceedings of the 8th Conference on Creativity and Cognition*, 81-90.

Megh Marathe and Priyank Chandra. 2020. Officers never type: examining the persistence of paper in e-governance. 2020. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 1-13.

Trevor H.J. Marchand. 2016. Introduction. Trevor H.J. Marchand (Ed.). *Craftwork as problem solving: ethnographic studies of design and making*. Farnham: Ashgate Publishing, 1-29.

Gary Marcus. 2018. Deep learning: a critical appraisal. *arXiv:1801.00631*. Accessed 15 December 2021.

Jean-Luc Marion. 2002. *Being given: toward a phenomenology of givenness*. Stanford, CA: Stanford University Press.

Noortje Marres. 2020. For a situational analytics: an interpretative methodology for the study of situations in computational settings. *Big Data & Society,* 7 (2), 1-16.

Alexander Mathews, Lexing Xie, and Xuming He. 2016. Senticap: generating image descriptions with sentiments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 30 (1), 2574-3580.

Eric Matthews 2002. *The philosophy of Merleau-Ponty*. Montreal: McGill-Queen's University Press.

John McCarthy. 2000. Approximate objects and approximate theories. In *Proceedings of the 7th International Conference on Principles of Knowledge Representation and Reasoning*, 519–526.

Malcolm McCullough. 1998. *Abstracting craft: the practiced digital hand*. Cambridge, MA: MIT Press.

David McGookin. 2019. Reveal: investigating proactive location-based reminiscing with personal digital photo repositories. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 1-14.

Jack Meadows. 2001. *Understanding information*. Munchen: K. G. Saur.

Nicolai Meinshausen. 2007. Relaxed lasso. *Computational Statistics & Data Analysis,* 52 (1), 374–393.

Maurice Merleau-Ponty. 1962. *The phenomenology of perception*. London: Routledge and Kegan Paul.

Johanna Meurer, Claudia Müller, Carla Simone, Ina Wagner, and Volker Wulf. 2018. Designing for sustainability: key issues of ICT projects for ageing at home. *Computer Supported Cooperative Work (CSCW),* 27 (3), 495-537.

Marius Mikalsen, Babak A. Farshchian, and Yngve Dahl. 2018. Infrastructuring as ambiguous repair: a case study of a surveillance infrastructure project. *Computer Supported Cooperative Work (CSCW),* 27 (2), 177-207.

Marius Mikalsen and Eric Monteiro. 2018. Data handling in knowledge infrastructures: a case study from oil exploration. In *Proceedings of the ACM on Human-Computer Interaction,* 2 (CSCW), 1-16.

Harvey J. Miller. 2010. The data avalanche is here. Shouldn’t we be digging? *Journal of Regional Science*, 50 (1), 181-201.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency,* 1-10.

Brent Mittelstadt. 2019, AI ethics–too principled to fail. *arXiv preprint arXiv:1906.06668*.

Annemarie Mol. 1999. Ontological politics. A word and some questions. *The Sociological Review,* 47 (1_suppl), 74-89.

Annemarie Mol. 2002. *The body multiple: ontology in medical practice*. Durham and London: Duke University Press.

Dyfed Rhys Morgan, Eifiona Thomas Lane, and David Styles. 2020. Crafty marketing: an evaluation of distinctive criteria for “craft” beer. *Food Reviews International*, 1-17.

Anders I. Mørch and Nikolay D. Mehandjiev. 2000. Tailoring as collaboration: the mediating role of multiple representations and application units. *Computer Supported Cooperative Work (CSCW*), 9 (1), 75-100.

Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. With most of it being pictures now, I rarely use it: understanding Twitter's evolving accessibility to blind users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 5506-5516.

Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-11.

Cecily Morrison, Edward Cutrell, Anupama Dhareshwar, Kevin Doherty, Anja Thieme, and Alex Taylor. 2017. Imagining artificial intelligence applications with people with visual disabilities using tactile ideation. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 81-90.

Michael Muller, Katja Neureiter, Nervo Verdezoto, Alina Krischkowsky, Anna Maria Al Zubaidi-Polli, and Manfred Tscheligi. 2016. Collaborative appropriation: how couples, teams, groups and communities adapt and adopt technologies. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 473-480.

Michael Muller, Melanie Feinberg, Timothy George, Steven J. Jackson, Bonnie E. John, Mary Beth Kery, and Samir Passi. 2019a. Human-centered study of data science work practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems,* 1-8.

Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019b. How data science workers work with data: discovery, capture, curation, design, creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-15.

Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, and Casey Dugan. 2021. Designing ground truth and the social life of labels. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-16.

Michael Muller and Angelika Strohmayer. 2022. Forgetting practices in the data sciences." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-19.

Brad A. Myers, Andrew J. Ko, Chris Scaffidi, Stephen Oney, YoungSeok Yoon, Kerry Chang, Mary Beth Kery and Toby Jia-Jun Li. 2017. Making end-user development more natural. In Fabio Paternò and Volker Wulf (Eds.), *New perspectives in end-user development*. Cham: Springer International Publishing, 1-22.

Ayman Nadeem. 2022. Human-centered approach to static-analysis-driven developer tools. *Communications of the ACM*, 65 (3), 38-45.

Nils J. Nilsson. 2007. 2007. The physical symbol system hypothesis: status and prospects. *50 Years of Artificial Intelligence*, 9-17.

Don A. Norman. 1988. *The design of everyday things*. New York: Doubleday.

William T. Odom, Abigail J. Sellen, Richard Banks, David S. Kirk, Tim Regan, Mark Selby, Jodi L. Forlizzi, and John Zimmerman. 2014. Designing for slowness, anticipation and re-visitation: a long term field study of the photobox. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1961-1970.

Dietmar Offenhuber. 2019. Data by proxy — material traces as autographic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 26 (1), 98-108.

Dietmar Offenhuber. 2020. What we talk about when we talk about data physicality. *IEEE Computer Graphics and Applications*, 40 (6), 25-37.

Bjørnar Olsen. 2010. *In defense of things: archaeology and the ontology of objects*. Lanham, MD: Rowman Altamira.

Wanda J. Orlikowski. 1992. The duality of technology: rethinking the concept of technology in organizations. *Organization Science*, 3 (3), 398-427.

Wanda J. Orlikowski and C. Suzanne Iacono. 2001. Desperately seeking the "IT" in IT research - a call to theorizing the IT artifact. *Information Systems Research*, 12 (2), 121-134.

Julian E. Orr. 1996. *Talking about machines: an ethnography of a modern job*. Ithaca, NY: Cornell University Press.

Cesc Chunseong Park, Byeongchang Kim, and Gunhee Kim. 2017. Attend to you: personalized image captioning with context sequence memory networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 895-903.

Soya Park, April Yi Wang, Ban Kawas, Q. Vera Liao, David Piorkowski, and Marina Danilevsky. 2021. Facilitating knowledge sharing from domain experts to data scientists for building NPL models. In *Proceedings of the International Conference on Intelligent User Interfaces*, 585-596.

Samir Passi and Steven J. Jackson. 2017. Data vision: learning to see through algorithmic abstraction. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2436-2447.

Samir Passi and Steven J. Jackson. 2018. Trust in data science: collaboration, translation, and accountability in corporate data science projects. In *Proceedings of the ACM on Human-Computer Interaction,* 2 (CSCW), 136, 1-28.

Samir Passi and Solon Barocas. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 39-48.

Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society*, 7 (2), 1-13.

Fabio Paternò and Volker Wulf (Eds.). 2017. *New perspectives in end-user development*. Cham: Springer International Publishing.

S. Tejaswi Peesapati, Victoria Schwanda, Johnathon Schultz, Matt Lepage, So-yae Jeong, and Dan Cosley. 2010. Pensieve: supporting everyday reminiscence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2027-2036.

Alex S. Pentland. 2013. The data-driven society. *Scientific American,* 309 (4), 78–83.

Daniela Petrelli and Steve Whittaker. 2010. Family memories in the home: contrasting physical and digital mementos. *Personal and Ubiquitous Computing*, 14, 153-169.

Helen Petrie, Chandra Harrison, and Sundeep Dev. 2005. Describing images on the web: a survey of current practice and prospects for the future. In *Proceedings of Human Computer Interaction International (HCII),* 71, 2.

Mikael Pettersson. 2011. Depictive traces: on the phenomenology of photography. *Journal of Aesthetics and Art Criticism*, 69 (2), 185-196.

Wolfgang Pietsch. 2022. *On the epistemology of data science: conceptual tools for a new inductivism*. Cham: Springer International Publishing.

Kathleen H. Pine and Max Liboiron. 2015. The politics of measurement and action. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 3147-3156.

Sarah Pink. 2011. Multimodality, multisensoriality and ethnographic knowing: social semiotics and the phenomenology of perception. *Qualitative Research*, 11 (3), 261-276.

David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. 2021. How AI developers overcome communication challenges in a multidisciplinary team: a case study. In *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW1), 131, 1-25.

Volkmar Pipek. 1995. *From tailoring to appropriation support: negotiating groupware usage*. Ph.D. Dissertation. Oulu University, Oulu, Finland. Available at http://herkules.oulu.fi/isbn9514276302. Accessed 20 December 2020.

Volkmar Pipek and Volker Wulf. 2009. Infrastructuring: toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems*, 10 (5), 447-473.

Volkmar Pipek, Helena Karasti, and Geoffrey C. Bowker. 2017. A preface to 'infrastructuring and collaborative design.' *Computer Supported Cooperative Work (CSCW)*, 26 (1), 1-5.

Neoklis Polyzotis, Sudip Roy. Steven Euijong Whang, and Martin Zinkevich. 2017. Data management challenges in production machine learning. In *Proceedings of the ACM International Conference on Management of Data, Chicago*, 1723-1726.

Theodore M. Porter. 1995. *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Manish Raghavan. 2021. *The societal impacts of algorithmic decision-making*. Doctoral dissertation, Cornell University.

Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 33-44.

Vijayshankar Raman and Joseph M. Hellerstein. 2001. Potter's wheel: an interactive data cleaning system. In *Proceedings of the 27th VLDB Conference*, 381-390.

Krishnan Ramnath, Simon Baker, Lucy Vanderwende, Motaz El-Saban, Sudipta Sinha, Anitha Kannan, Noran Hassan, Michel Galley, Yi Yang, Deva Ramanan, Alessandro Bergamo, and Lorrenzo Torresani. 2014. AutoCaption: automatic caption generation for personal photos. *IEEE Winter Conference on Applications of Computer Vision*, 1050-1057.

Matt Ratto. 2011. Critical making: conceptual and material studies in technology and social life. *The Information Society*, 27 (4), 252-260.

Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi, and Meredith Ringel Morris. 2017. Eyes-free art: exploring proxemic audio interfaces for blind and low vision art engagement. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1 (3), 1-21.

Thomas C. Redman. 2018. If your data is bad, your machine learning tools are useless. *Harvard Business Review.* https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless. Accessed 15 December 2021.

Johan Redström and Heather Wiltse. 2018. *Changing things: the future of objects in a digital world*. London: Bloomsbury Publishing.

David Ribes. 2019. STS, meet data science, once again. *Science, Technology, & Human Values*, 44 (3), 514-539.

David Ribes, Andrew S. Hoffman, Steven C. Slota, and Geoffrey C. Bower. 2019. The logic of domains. *Social Studies of Science*, 49 (3), 281-309.

David Ribes and Steven J. Jackson. 2013. Data bite man: the work of sustaining a long-term study. In Lisa Gitelman (Ed.), *Raw data is an oxymoron*. Cambridge, MA: MIT Press, 147-166.

Yuji Roh, Geon Heo, and Steven Euijong Whang. 2019. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33 (4), 1328-1347.

Daniel Rosenberg. 2013. Data before the fact. In Lisa Gitelman (Ed.), *Raw data is an oxymoron*. Cambridge, MA: MIP Press, 15-40.

Robert Rosenberger. 2018. Why it takes both postphenomenology and STS to account for technological mediation. In Jesper Aagaard, Jan Kyrre Berg Friis, Jessica Sorenson, Oliver Tafdrup, and Catherine Hasse (Eds.)., *Postphenomenological methodologies.* Lanham, MD: Lexington Books, 171-198.

Daniela K. Rosner. 2012. The material practices of collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 155-1164.

Daniela K. Rosner and Morgan Ames. 2014. Designing for repair? Infrastructures and materialities of breakdown. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 319-331.

Royal Society. 2012. *Science as an open enterprise*. http://royalsociety.org/policy/projects/science-public-enterprise/report/.

Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1 (5), 206-215.

Cynthia Rudin and Joanna Radin. 2019. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review,* 1 (2).

Mauricio Sáenz and Jaime Sánchez. 2009. Indoor position and orientation for the blind. *International Conference on Universal Access in Human-Computer Interaction*, 236-245.

Elliot Salisbury, Ece Kamar, and Meredith Morris. 2017. Toward scalable social alt text: conversational crowdsourcing as a tool for refining vision-to-language technology for the blind. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 5 (1), 147-156.

Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Kumar Paritosh, and Lora Mois Aroyo. 2021. "Everyone wants to do the model work, not the data work": data cascades in high-stakes AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-15.

Nithya Sambasivan and Rajesh Veeraraghavan. 2022. The deskilling of domain expertise in AI development. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.

Pedro Sanches and Barry Brown. 2018. Data bites man: the production of malaria by technology. In *Proceedings of the ACM on Human-Computer Interaction*, 2 (CSCW), 153, 1-19.

Iqbal H. Sarker. 2021. Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, 2 (3), 1-21.

Risto Sarvas and David M. Frohlich. 2011. *From snapshots to social media: the changing picture of domestic photography*. London: Springer.

Anastasia Schaadhardt, Alexis Hiniker, and Jacob O. Wobbrock. 2021. Understanding blind screen-reader users' experiences of digital artboards. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-19.

Morgan K. Scheuerman, Hanna Alex, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. In *Proceedings of the ACM on Human-Computer Interaction,* 5 (CSCW2), 317, 1-37.

Kjeld Schmidt and Carla Simone. 1996. Coordination mechanisms: towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)*, 5 (2-3), 155-200.

Kjeld Schmidt and Ina Wagner. 2004. Ordering systems: coordinative practices and artifacts in architectural design and planning. *Computer Supported Cooperative Work (CSCW),* 13 (5-6), 349-408.

Johannes Schobel, Rüdiger Pryss, Winfried Schlee, Thomas Probst, Dominic Gebhardt, Marc Schickler, and Manfred Reichert. 2017. Development of mobile data collection applications by domain experts: experimental results from a usability study. In *Proceedings of the International Conference on Advanced Information Systems Engineering*, 60-75.

Cornelius Schubert. 2019. Repair work as inquiry and improvisation: the curious case of medical practice. Ignaz Strebel, Alain Bovet, and Philippe Sormani (Eds.). *Repair work ethnographies*. Singapore: Palgrave Macmillan, 129-160.

David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. 2015. Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28, 2503-2511.

Seeing AI. https://www.microsoft.com/en-us/ai/seeing-ai.

Judith Segal. 2009. Software development cultures and cooperation problems: a field study of the early stages of development of software for a scientific community. *Computer Supported Cooperative Work (CSCW),* 18 (5), 581-606.

Catherine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Co-designing data experiments: domain experts' exploration and experimentation with self-selected data sources. In *Proceedings of the Nordic Conference on Human-Computer Interaction,* 1-11.

Abigail J. Sellen and Richard H.R. Harper. 2003. *The myth of the paperless office*. Cambridge, MA: MIT Press.

Abigail J. Sellen and Steve Whittaker. 2010. Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM,* 53 (5), 70-77.

Richard Sennett. 2008. *The craftsman*. New Haven CT: Yale University Press.

Ji-Eun Seo. 2018. Maker of Cass buys local craft beer company. *JoongAng Daily*. https://koreajoongangdaily.joins.com/2018/04/04/industry/Maker-of-Cass-buys-local-craft-beer-company/3046509.html. Accessed 10 May 2021.

Michel Serres.1996. *Genesis*. Geneviève James and James Nielson (Trans.). Ann Arbor, MI: University of Michigan Press.

Tom Seymoens, Femke Ongenae, An Jacobs, Stijn Verstichel, and Ann Ackaert. 2018. A methodology to involve domain experts and machine learning techniques in the design of human-centered algorithms. *IFIP Working Conference on Human Work Interaction Design,* 200-214.

Nishant Shah. 2013. Big data, people's lives, and the importance of openness. *Connected Learning Alliance.* https://clalliance.org/blog/big-data-people-s-lives-and-the-importance-of-openness/.

Galit Shmueli. 2010. To explain or to predict? *Statistical Science,* 25 (3), 289-310.

Ben Shneiderman. 2020. Human-centered artificial intelligence: three fresh ideas. *AIS Transactions on Human-Computer Interaction,* 12 (3), 109–124.

Ben Shneiderman. 2022. *Human-centered AI*. Oxford: Oxford University Press.

Kurt Shuster, Samuel Humeau, Hexiang Hu, Antoine Bordes, and Jason Weston. 2019. Engaging image captioning via personality. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 12516-12526.

Rachel N. Simons, Danna Gurari, and Kenneth R. Fleischmann. 2020. "I hope this is helpful" Understanding crowdworkers' challenges and motivations for an image description task. In *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW2), 1-26.

Benjamin Sims and Christopher R. Henke. 2012. Repairing credibility: repositioning nuclear weapons knowledge after the Cold War. *Social Studies of Science,* 42 (3), 324-347.

Ronal Singh, Paul Dourish, Piers Howe, Tim Miller, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. 2021. Directive explanations for actionable explainability in machine learning applications. *arXiv preprint arXiv:2102.02671*.

Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. Participation is not a design fix for machine learning. *arXiv preprint arXiv:2007.02423*.

Brain Cantwell Smith. 1998. *On the origin of objects*. Cambridge, MA: MIT Press.

Brian Cantwell Smith. 2019. *The promise of artificial intelligence: reckoning and judgment.* Cambridge, MA: MIT Press.

Susan Sontag. 1977. *On photography*. New York, NY: Picador.

Elizabeth V. Spelman. 2012. *Repair: the impulse to restore in a fragile world*. Boston: Beacon Press.

Thilo Stadelmann, Mohammadreza Amirian, Ismail Arabaci, Marek Arnold, Gilbert François Duivesteijn, Ismail Elezi, Melanie Geiger, Stefan Lörwald, Benjamin Bruno Meier, Katharina Rombach, and Lukas Tuggener. 2018. Deep learning in the wild. *IAPR Workshop on Artificial Neural Networks in Pattern Recognition,* 17-38.

Abigale J. Stangl, Esha Kothari, Suyog D. Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. BrowseWithMe: an online clothes shopping assistant for people with visual impairments. In

Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility, 107–118.

Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, shoes, tree. Is the person naked?" What people with vision impairments want in image descriptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-13.

Abigale Stangl, Nitin Verma, Kenneth R. Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 1-15.

Susan Leigh Star and Karen Ruhleder. 1996. Steps toward an ecology of infrastructure: design and access for large information spaces. *Information Systems Research,* 7 (1), 111-134.

Susan Leigh Star. 2010. This is not a boundary object: reflections on the origin of a concept. *Science, Technology & Human Values,* 35 (5), 601-617.

Michael Stonebraker. 1989. Inclusion of new types in relational database systems. In John Mylopolous and Michael Brodie (Eds.), *Readings in Artificial Intelligence and Databases*, San Mateo, CA: Morgan Kaufmann, 599-606.

Lucy A. Suchman. 1983. Office procedure as practical action: models of work and system design. *ACM Transactions on Information Systems (TOIS),* 1 (4), 320-328.

Lucy A. Suchman. 2002. Located accountabilities in technology production. *Scandinavian Journal of Information Systems,* 14 (2), 91-105.

Lucy A. Suchman 2005. Affiliative objects. *Organization,* 12 (3), 379-399.

Lucy A. Suchman. 2007. *Human-machine reconfigurations: plans and situated actions.* New York: Cambridge University Press.

Sullivan+. https://blog.mysullivan.org/search/label/Sullivan_Plus_20Function_Guide.

Sharifa Sultana, Md Mobaydul Haque Mozumder, and Syed Ishtiaque Ahmed. 2021. Chasing luck: data-driven prediction, faith, hunch, and cultural norms in rural betting practices. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-17.

Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. 2017. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision*, 843-852.

Saiganesh Swaminathan, Thijs Roumen, Robert Kovacs, David Stangl, Stefanie Mueller, and Patrick Baudisch. 2016. Linespace: a sensemaking platform for the blind. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2175-2185.

David Taieb. 2018. *Thoughtful data science*. Birmingham, UK: Packt Publishing.

TalkBack. https://support.google.com/accessibility/android/topic/10601571?hl=en&ref_topic=3529932.

Anissa Tanweer, Emily Kalah Gade, P.M. Krafft, and Sarah K. Dreier. 2021. Why the data revolution needs qualitative thinking. *Harvard Data Science Review,* 3 (3).

TapTapSee. https://taptapseeapp.com.

Alex S. Taylor, Siân Lindley, Tim Regan, David Sweeney, Vasillis Vlachokyriakos, Lillie Grainger, and Jessica Lingel. 2015. Data-in-place: thinking through the relations between data and community. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 2863-2872.

Andrea K. Thomer, Michael Bernand Twidale, and Matthew J. Yoder. 2019. Transforming taxonomic interfaces: "arm's length" cooperative work and the maintenance of a long-lied classification system. In *Proceedings of the ACM on Human-Computer Interaction,* 2 (CSCW), 1-23.

Nigel Thrift. 2008. *Non-representational theory: space, politics, affect*. London: Routledge.

Vasiliki Tsaknaki, Marisa Cohn, Laurens Boer, Ylva Fernaeus, and Anna Vallgårda. 2016. Things fall apart: unpacking the temporalities of impermanence for HCI. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 1-3.

Manfred Tscheligi, Alina Krischkowsky, Katja Neureiter, Kori Inkpen, Michael Muller, and Gunnar Stevens. 2014. Potentials of the "unexpected" technology appropriation practices and communication needs. In *Proceedings of the 18th International Conference on Supporting Group Work*, 313-316.

Edward R. Tufte. 2001. *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

John W. Tukey and Martin B. Wilk. 1966. Data analysis and statistics: an expository overview. In *Proceedings of the Fall Joint Computer Conference,* 695-709.

Daniel Tunkelang. 2009. Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services,* 1 (1), 1-80.

Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable recourse in linear classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency,* 10–19.

Nancy Van House. 2016. Photographic wayfaring, now and to come. In Edgar Gomez Cruz and Asko Lehmuskallio (Eds.), *Digital photography and everyday life: empirical studies on material visual practices*. London and New York: Routledge, 274-283.

Max Van Manen. 2007. Phenomenology of practice. *Phenomenology & Practice,* 1 (1), 11-30.

Max Van Manen. 2014. *Phenomenology of practice: meaning-giving methods in phenomenological research and writing*. New York and London: Routledge.

Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*, 95-102.

Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.

Briana Vecchione, Solon Barocas, and Karen Levy. 2021. Algorithmic auditing and social justice: lessons from the history of audit studies. In *Proceedings of the Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1-9.

Peter-Paul Verbeek. 2005. *What things do: philosophical reflections on technology, agency, and design.* University Park, PA: Pennsylvania State University Press.

Sahil Verma, John Dickerson, and Keegan Hines. 2020. Counterfactual explanations for machine learning: A review. *arXiv preprint arXiv:2010.10596*.

Stijn Viaene. 2013. Data scientists aren't domain experts. *IT Professional*, 15 (6), 12-17.

Beatrice Vincenzi, Alex S. Taylor, and Simone Stumpf. 2021. Interdependence in action: people with visual impairments and their guides co-constituting common spaces. In *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1), 1-33.

VoiceOver Recognition. https://support.apple.com/en-us/HT211899.

Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How blind people interact with visual content on social networking services. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 1584-1595.

Kiri Wagstaff. 2012 Machine learning that matters. *arXiv preprint arXiv:1206.4656*. Accessed 11 January 2022.

April Yi Wang, Anant Mittal, Christopher Brooks, and Steve Oney. 2019a. How data scientists use computational notebooks for real-time collaboration. In *Proceedings of the ACM on Human-Computer Interaction,* 3 (CSCW), 39, 1-30.

Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019b. Human-AI collaboration in data science: exploring data scientists' perceptions of automated AI. In *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), 1-24.

Dakuo Wang, Josh Andres, Justin D. Weisz, Erick Oduor, and Casey Dugan. 2021. AutoDS: towards human-centered automation of data science. In *Proceedings of the CHI Conference on Human Factors in Computing Systems,*1-12.

Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019c. Designing theory-driven user-centric explainable AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-15.

Ruolin Wang, Zixuan Chen, Mingrui Ray Zhang, Zhaoheng Li, Zhixiu Liu, Zihan Dang, Chun Yu, and Xiang Anthony Chen. 2021. Revamp: enhancing accessible information seeking experience of online shopping for blind or low vision users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.

Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31 (2), 841-888.

Jeffrey Dean Webster and Mary E. McCall. 1999. Reminiscence functions across adulthood: a replication and extension. *Journal of Adult Development*, 6 (1), 73-85.

Adrian Weller. 2019. Transparency: motivations and challenges. In Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, Klaus-Robert Muller (Eds.), *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer, 23–40.

Steve Whittaker, Ofer Bergman, and Paul Clough. 2010. Easy on that trigger dad: a study of long term family photo retrieval. *Personal and Ubiquitous Computing*, 14 (1), 31-43.

Maranke Wieringa. 2020. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 1-18.

Sinead A. Williamson and Jette Henderson. 2021. Understanding collections of related datasets using dependent MMD coresets. *Information*, 12 (10), 392.

Christine Wolf, Kathryn E. Ringland, and Gillian Hayes. 2019. Home worlds: situating domestic computing in everyday life through a study of DIY home repair. In *Proceedings of the ACM on Human-Computer Interaction*, 3 (CSCW), 1-22.

Qi Wu, Peng Wang, Chunhua Shen, Anthony Dick, and Anton van den Hengel. 2016. Ask me anything: free-form visual question answering based on knowledge from external sources. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4622–4630.

Qi Wu, Damien Teney, Peng Wang, Chunhua Shen, Anthony Dick, and Anton van den Hengel. 2017. Visual question answering: a survey of methods and datasets. *Computer Vision and Image Understanding*, 163, 21-40.

Shaomei Wu, and Lada A. Adamic. 2014. Visually impaired users on an online social network. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 3133-3142.

Susan Wyche, Tawanna R. Dillahunt, Nightingale Simiyu, and Sharon Alaka. 2015. "If god gives me the chance I will design my own phone": exploring mobile phone repair and postcolonial approaches to design in rural Kenya. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 463-473.

Doris Xin, Eva Yiwei Wu, Doris Jung-Lin Lee, Niloufar Salehi, and Aditya Parameswaran. 2021. Whither AutoML? Understanding the role of automation in machine learning workflows. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-16.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: neural image caption generation with visual attention. *International Conference on Machine Learning*, 2048-2057. Available from https://proceedings.mlr.press/v37/xuc15.html.

MinYoung Yoo, William Odom, and Arne Berger. 2020. Understanding how audio mediates experiences of reminiscence for people living with blindness. In *Companion Publication of the ACM Designing Interactive Systems Conference*, 73-78.

MinYoung Yoo, William Odom, and Arne Berger. 2021. Understanding everyday experiences of reminiscence for people with blindness: practices, tensions and probing new design possibilities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-15.

Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A. Miele, and Ilmi Yoon. 2020. Human-in-the-loop machine learning to increase video accessibility for visually impaired and blind users. In *Proceedings of the ACM Designing Interactive Systems Conference*, 47-60.

Xiaoyu Zeng, Yanan Wang, Tai-Yin Chiu, Nilavra Bhattacharya, and Danna Gurari. 2020. Vision skills needed to answer visual questions. In *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW2), 149, 1-31.

Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The effect of computer-generated descriptions on photo-sharing experiences of people with visual impairments. In *Proceedings of the ACM on Human-Computer Interaction*, 1 (CSCW), 121, 1-22.

Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: understanding use beyond the lab. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.

Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How do data science workers collaborate? Roles, workflows, and tools. In *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW1), 22, 1-23.

Yu Zhong, Walter S. Lasecki, Erin Brady, and Jeffrey P. Bigham. 2015. RegionSpeak: quick comprehensive spatial descriptions of complex images for blind users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2353–2362.