CHAPTER

Online and ubiquitous HCI research 14

14.1 INTRODUCTION

Where and how do humans interact with computers? Much of the early work in human-computer interaction (HCI) research—and, indeed, some of the content in this book—focused on traditional computers—monitors and keyboards on desks, perhaps with mice or other input devices by their side. However, that is far from the whole story. As important as traditional computing has been and continues to be, much of how we interact with computers has moved from beyond the desktop onto the Internet and beyond. Social media, crowdsourcing, connected devices, and the "Internet of Things" all present interesting opportunities across the spectrum of human-computer interaction research—from understanding needs to evaluating systems and then studying how those systems are used.

This chapter attempts to tie together areas of work that might at first seem disjoint. *Online research* discusses techniques for conducting remote usability studies and other internet-enabled research, including studies of social media and online communities (online surveys are covered in Chapter 5). *Human computation* discusses the use of online tools that ask large numbers of users to perform small tasks—an approach that has proven very useful for many HCI studies. *Sensors and Ubiquitous computing* expands upon the cell phones and fitness monitoring devices described in Chapter 13, to include the widespread use of inexpensive sensors to measure aspects of the world around us, providing augmented depictions of daily life and everyday environments.

Although these topics may seem very different, they share the common thread of investigating computer use outside of traditional contexts and goals. Online studies and ubiquitous computing research investigate the role of computing in social and everyday environments that would not have been possible in the early days of HCI research in the 1980s. Similarly, human computation studies envision novel approaches of the power of connected communities of people to solve otherwise difficult problems. We will discuss some examples of these new forms of computing, and how they might inform and extend the possibilities of your HCI research.

Research Methods in Human-Computer Interaction. http://dx.doi.org/10.1016/B978-0-12-805390-4.00014-5

14.2 ONLINE RESEARCH

Finding the right way to engage and interact with potential participants is a constant struggle for HCI researchers. Offering a small cash payment and perhaps some free food to nearby undergraduates may help to some extent, but such convenience samples raise concerns about validity and generalizability. Finding users who may represent a broad range of ages, skills, and backgrounds may require getting out of the lab to engage with a broader range of participants. These challenges become even more profound for studies requiring specialized populations, such as highly trained domain specialists or users with particular disabilities, who might be both hard to identify and hard to get to the lab, due to their busy schedules and other constraints. Like many of our colleagues, we all have had the experience of struggling to find times to meet with those hard-to-engage participants, traveling around our cities to conduct studies in participants' homes, and otherwise worrying about our ability to find the right folks to finish our studies.

Although certainly no panacea, the Internet can help. Beyond the obvious approach of using message boards and mailing lists to recruit participants, HCI researchers have found various types of online studies to be appealing and effective. Although not without their pitfalls, online research studies can, when designed correctly, help HCI researchers extend their reach and complete studies with less difficulty and expense. The unique challenges, pitfalls, and opportunities of online research should be considered carefully before starting any studies.

Moving beyond simply conducting traditional studies online, this section also looks at online activity as the focus of HCI research studies. These studies—which are inherently "online research"—explore the dynamics of social interactions conducted online to understand how message boards, social media, and other tools enable social interaction and the spread of ideas.

14.2.1 OBSERVATIONAL ONLINE STUDIES

The classic HCI investigation involves watching a participant as they use a computer. Contextual inquiries; think-aloud and other usability studies; and empirical comparisons certainly differ in their design, content, and execution, but they share a common core: a participant in the study sits at the computer using a tool to complete a task, while one or more researchers watch, take notes, and record data. Being in the same room provides many advantages, as researchers are able to establish rapport with participants and build trust necessary for constructive conversations. Physical co-location also helps observant researchers learn from watching their participants, noting body language that might suggest discomfort, impatience, or other reactions to the tool or task at hand. Despite these benefits, in-person studies are also inherently limiting, as two (or more) people must find their way to the same location.

Noting these difficulties, HCI researchers have developed strategies for using Internet technologies to conduct usability studies remotely. Although web-based conferencing tools (also discussed in Section 12.3.3) may still seem relatively novel—and are still often difficult to use—the use of these tools for usability studies is at least 20 years old (Hartson et al., 1996). Web conferencing tools capable of remote screen-sharing (either one window at a time or full-screen) and integrated audio via voice-over IP or telephone, provide a basis for conversations between researcher and participant, both looking at the same screen content. Some tools go further, providing webcam video for participants, recording capabilities, and even remote mouse/keyboard control, providing one participant in the conversation to (with permission) control their interlocutor's computer.

Given these features, it is quite possible to conduct contextual inquiry and think-aloud studies online. Contextual inquiries are perhaps easiest: your participant can share his or her screen, start their work, and you can sit back and watch, asking questions as needed. The click-by-click view of the tool at hand will provide a detailed picture of how individual actions are taken to complete tasks, and answers to questions should help you understand the work. Think-aloud studies that use traditional desktop software are slightly more challenging, as your participants will probably not have your software installed on their computers. Thus, web applications are particularly well suited for remote think-aloud studies, as users can access your web site from their computer, just as they would if the tool were deployed on a production server. Alternatively, you might be able to send your participants an installable version of the software that they might run locally. If this is not possible, remote mouse and keyboard capabilities might help users control the software on your machine. In any case, once these details are worked out the study can proceed much like any other think-aloud.

Choosing the correct tools for this sort of work is critical. Beyond basic screen-sharing and integrated audio, functionality for recording sessions is invaluable. The ability to replay and review sessions will augment your memory and free you up to focus on the participants' comments and actions, rather than on note-taking. Webcam video can be a great way to see the participants' facial expression and body language, restoring some of the fidelity lost when the participant is not in the room with the researcher. However, this value is limited—many tools provide only webcam or screen sharing, but not both simultaneously. Remote mouse keyboard/control can be very helpful, but only if it works well. If you are planning on using this remote control facility, do not forget that your computer will be unavailable to you while the participant is controlling things—you will not be able to use your own machine to take notes. A second computer might be needed. Do your due diligence before relying on one of these tools to conduct your study—compare features, use free trials to see how well services really work, and try multiple pilot sessions.

Other technical challenges associated with online observational studies include firewalls and desktop configurations. Some users may work in institutions that might be unfriendly to the use of these tools, potentially blocking access. Network policies at your own institution may be a difficulty as well, as firewall exceptions might be needed to place your web application on a system that is visible to the outside world. Web conferencing tools might require the installation of plugins on the participants' computers, a process that might be difficult or impossible in some working environments. For these reasons, you might also ask each participant to conduct a "test-drive" before your main session. If you only have a limited amount of time for each participant, you do not want to spend too much of it working on software configuration.

Even if all of the technology works perfectly, you should be prepared for a different experience when conducting online observational studies. The lack of physical presence will make it harder to gauge participants' reactions, and even to know how intently they are focusing on the task (Dray and Siegel, 2004). More bluntly, you might have no way of knowing when users are surfing the web as opposed to attending to your questions. You might also find that contextual inquiries are somewhat limited: webcams might do a great job of showing faces, but they will not help you see anything that does not involve computer work, such as filing papers or reviewing printed material. Finally, do not forget that any plans to record sessions should be approved by your institutional review board or equivalent (see Chapter 15).

Online studies are not limited simply to web-conferencing systems. HCI researchers have experimented with other techniques designed to address shortcomings of both in-person and online studies. One effort found that 3D virtual world simulation of a usability lab provided some advantages over a web-conferencing-based usability study (Madathil and Greenstein, 2011). Webcam-based eye trackers (see Chapter 13) have also been used to remotely collect low-level interaction data as needed for usability studies (Chynał and Szymański, 2011). Alternatively, you might consider simpler, more low-tech approaches. Online reporting of critical usability incidents and posting of usability problems to online forums have been shown to be effective for identifying usability problems, although at a lower rate than in-person usability tests (Andreasen et al., 2007; Bruun et al., 2009). Such studies also have the advantage of being potentially asynchronous—you might ask participants to complete tasks at their convenience, reporting usability problems as appropriate. As is often the case with usability studies and expert reviews, providing specific predefined tasks may help participants identify more usability problems (Bruun and Stage, 2012).

Remote online usability studies can be useful for recruiting and including hard-to-find participants, such as individuals with disabilities who might have some difficulty in making the trip to a usability lab. Although this approach has been shown to have some potential utility, technologies should be chosen carefully to fit the needs of target populations (Petrie et al., 2006) (see Chapter 16 for further discussion of HCI research involving people with disabilities). Remote studies also provide for the possibility of software enhancements for people with specific disabilities, such as sign-language facilities for use by deaf participants (Schnepp and Shiver, 2011).

14.2.2 ONLINE DATA COLLECTION

Online HCI techniques are not limited to usability and think-aloud studies. Online surveys have become very familiar—see Chapter 5 for a discussion of the ins and outs of conducting online surveys. Chapter 12 discusses two useful online research techniques: the use of instrumented software to collect user interaction data and the use of web logs to study how web sites are used.

Web log analysis can be particularly useful for comparison of alternative web site designs or interactions. "A/B" testing is a widely used approach for comparing alternative designs for active web sites. In an A/B test a server is configured to randomly select one of two alternatives—the "A" and "B" designs to be presented whenever a visitor comes to a site. Given enough visits, data can be collected to see which users complete specified tasks more quickly or with fewer errors. Such tests might also add quick surveys asking users for their impression of a site. By using functioning web sites to gather data on many users who had come to a site, these A/B tests enable rapid collection of usability data, without the need to conduct a formal usability test.

Building on this approach, it is also possible to conduct empirical studies online. Just as web logs might be used to extract event and therefore task completion times in web-based studies run on a local server, an appropriately structured site might enable easy extraction of task completion times, results, etc. You can also create appropriate components of the web site to collect informed consent (with approval of your IRB, see Chapter 15), demographic information, and other needed details. Such an installation has the advantage of allowing participants to enroll in a study without your participation—they can just go to the URL in question and follow the directions.

Any timing data collected from either A/B testing or online empirical studies runs the risk of being confounded by network latencies or problems. If a network problem slows the communication between participants' computers and your servers, task completion times may be slowed, but you would not have any way of knowing that that had happened. Larger numbers of participants might help with this problem, as extreme values in latency will be more clearly identified as outliers.

Validity of online versus lab-based studies may be a concern. One study of the utility of online versus lab-based studies for empirical evaluation of search interfaces found that online and lab-based studies produced comparable results (Kelly and Gyllstrom, 2011). To ensure similar generalization to your problems of interest, you might consider pairing a small in-person study with a larger online study. Similarities in the results will increase confidence in the online data, but discrepancies might indicate some difficulties in translation (Meyer and Bederson, 1998).

A/B testing has been used extensively by companies with prominent Internet business activity, as Amazon, Microsoft, and other familiar web companies are well aware of the importance in small changes in design and task completion. For sites serving millions of users, an increase of even 1% on ad views or completed sales can mean significant increases in revenue. The importance of A/B testing has led to significant methodological interest, from practical guidance from web usability guru Jakob Nielsen (Nielsen, 2005, 2012, 2014) to papers on the design of A/B studies (Kharitonov et al., 2015) and the investigation of novel statistical analysis techniques (Deng et al., 2013, 2014; Deng, 2015). Ron Kohavi and colleagues at Microsoft have published extensively in this area, including a survey presenting a broad overview of the topic (Kohavi et al., 2009) and papers discussing some of the pitfalls and lessons learned from Microsoft's extensive A/B testing (Crook et al., 2009; Kohavi et al., 2012, 2013).

A/B testing is also limited by the coarse-grain nature of the data. Knowing which elements are clicked on which pages can be useful, but additional data might be needed to know where and how those pages command user attention. Eye-tracking techniques (Chapter 13) and additional software tools such as proxies and JavaScript libraries (Chapter 12) can provide finer-grain detail when necessary.

14.2.3 ONLINE ACTIVITY

The rich stores of data created through our online lives provide tantalizing HCI research possibilities. Exploration of online content and activity can provide deep insight into how people communicate, create communities, learn, and interact online, including how ideas develop and spread, and what we might learn from information dissemination patterns. Although the techniques are very similar to others discussed earlier in this book—including both qualitative content analysis (Chapter 11) and statistical review of automatically captured interaction data and human physiological signals (Chapters 12 and 13)—the domain is qualitatively different, in that analysis of online activity effectively involves the emergence of community and collective behavior.

14.2.3.1 Online communities

Computers have been used for online communities since the early 1980s, with the early USENET discussion groups on the ARPANET (Leug and Fisher, 2003) leading to online bulletin boards where home computer users with dial-up modems could interact. The growth of the Internet in the 1990s led to the emergence of countless bulletin boards for communities of interest, providing researchers with an opportunity to study communication and patterns, community growth, and related dynamics.

Analyses of these communities often combine qualitative and quantitative methods. Qualitative methods might include thematic content analysis (see Chapter 11), aimed at extracting common themes and types of interactions, perhaps guided by some theory. These studies will typically involve reading through large numbers of posts, coding contents for types of concerns, types of posts (questions, answers, guidance, emotional support), and for conversational structure (introduction of new members, arguments over controversial topics, resolutions of disputes, etc.). Although time-consuming, these techniques offer the possibility of immersion in the community under consideration, providing rich context that might enable deep understanding.

As online community content is often, if not exclusively, found in the form of online text, it is particularly well suited for automated analysis and quantitative investigation of patterns of interest, including how and when certain terms or types of discourse are used. Forum content and posts can generally be downloaded, although with varying levels of difficulty, depending on the underlying software platform. Communities built on open platforms might provide programming libraries known as Application Programming Interfaces (APIs) capable of extracting data. Using these libraries, software developers might develop custom programs to gather and collate data needed to address research questions of interest. Barring such facilities, researchers might need to resort to parsing HTML content. Once message content is available, it can be analyzed via textual analysis approaches, including building distributions of words or phrases used in different contexts. These data might then be stored in a database and analyzed for frequency of occurrence, sequences, or other patterns of interest. For a slightly more nuanced approach, natural-language processing techniques might be applied to posts to distinguish between, for example, cases where someone is discussing their own current concerns as opposed to those faced in the past, or experienced by a family member (Harkema et al., 2009). As with other triangulation approaches, these techniques might work best hand in hand, with qualitative insights suggesting patterns that might be quantified and quantitative identification of frequent behaviors driving new theories for qualitative exploration.

Diane Maloney-Krichmar and Jenny Preece's in-depth study of an online forum for people with knee injuries provides a rich example of the use of both qualitative and quantitative methods to understand the dynamics of a complex online community. Using a four-phase research plan, Maloney-Krichmar and Preece combined preliminary observation with usability analysis, detailed quantitative analysis of 1 week's worth of messages, and interviews with members of the site. Results included characterization of site features supporting sociability; membership patterns encouraging the health of the community, including identifiable subgroups; task and individual roles assumed by members; distribution of discussion length, including the number of messages in each thread; and characterization of the role of group participation in members' lives (Maloney-Krichmar and Preece, 2005). This detailed picture provides an example of the possibilities of applying ethnographic techniques to online communities.

HCI researchers have studied a wide range of online communities. Analyses of content and interviews with participants were used to develop detailed descriptions of a 2015 "protest" in the Reddit online community, during which volunteer moderators protested changes in company policy and staff (Centivany and Glushko, 2016; Matias, 2016). A study of contributions to a repository of projects developed using the online programming tool Scratch used review of published user profiles and comments on projects to explore the diversity of participants in the community (Richard and Kafai, 2016).

Online communities can be useful research resources even if you are not willing (or able) to undertake a detailed ethnographic analysis of a specific group's dynamics. As shared resources for individuals with common interests, these communities can often be valuable tools for recruiting participants for studies of factors surrounding community goals. One study of the dynamics of conflict in free and open-source software development conducted a survey involving participants in the software development site GitHub, home to many open-source projects (Filippova and Cho, 2016). Other studies might involve communities spanning multiple sites. A study of the credibility of medical "crowdfunding" requests (using online sites to solicit contributions to offset medical expenses) examined Reddit discussions regarding campaigns posted on other sites. Like other studies discussed earlier, this investigation combined content analysis of postings with participant interviews (Kim et al., 2016a).

14.2.3.2 Following trends: Social media and online interaction data

What can we learn from behavior on the online sites that seem to occupy so much of our collective attention? Moving beyond the closed confines of online communities, broader studies of both content and patterns of online activity can tell us a great deal about how people interact, how ideas spread, and what meaning might be attributed to those patterns.

Studies of online activity can be classified into three broad categories, distinguished by data source. *Social media* studies explore participation in familiar sites such as Twitter and Facebook to understand how these tools can be used to find and share information. In this context, we use the term "social media" to refer to general-purpose sites supporting individually selected lists of "friends" or "contacts," as opposed to interest-specific communities described in Section 14.2.3.1. Examples include studies of how people use social media to meet information needs (Menefee et al., 2016), and examinations of the impact of social media on dissemination of information from research conferences (Winandy et al., 2016). *Web search* studies examine queries submitted to general Internet search engines, looking for behaviors common to many web users, such as searching for information about flu outbreaks (Ginsberg et al., 2009) and other health conditions (White et al., 2013; Paparrizos et al., 2016). Examinations of *blogs*, *wikis*, *and other user-generated content* explore how users interact in creating and sharing information on the web, including video blogs (Huh et al., 2014), Wikipedia editing (Viégas et al., 2004, 2007a,b; Kittur and Kraut, 2008), and online reviews (Hedegaard and Simonsen, 2013, 2014), to name a few. Boundaries between these categories are fuzzy, and many of these goals can be met by multiple sources of interaction data.

Identification of appropriate data sources, and of the means of accessing that data, is often the first step in conducting studies of online interactions. Designing a study to investigate the use of "social media" in examining a topic of interest is a reasonable start, but details are important—which social media sites will you consider? Which content types? Various sources will differ significantly in their willingness to share data and in the tools available to access any data that is openly available. Open-source sites like Wikipedia might allow access to data that might be considered proprietary by for-profit search engines. Some social media sites such as Facebook (https://developers.facebook.com/docs/graph-api) and Twitter (https://dev.twitter.com/overview/documentation) sites provide API access suitable for querying data sets, while others may require the use of more manual tools to "screen-scrape" data off of web pages. However, the mere presence of an API might not be sufficient—APIs that limit the quantity or range of content that can be retrieved might not be sufficient for some tasks.

An examination of selected papers provides a sampling of some of the approaches researchers have used to access social media interaction data. Small-scale studies—such as examining the impact and diffusion of social media content for a specific issue among a small community—can be conducted relatively easily. Organizers of a 2011 health research conference established social media presences on Facebook, Twitter, Flickr, and other sites and tracked the utilization and dissemination of content over time as a means of examining the impact of their efforts (Winandy et al., 2016). Such focused efforts have the advantage of generally being feasible with information available to account holders on these sites. Other, similarly small studies, can be conducted through standard interactions, as in a study of YouTube video blogs for illness support: researchers manually searched YouTube to identify videos of interest and reviewed transcripts and comments on those videos to see how they were used for social support (Huh et al., 2014)

For larger studies, APIs provided by vendors are often the most effective means of capturing data. Twitter APIs have been used to access data for many studies, including investigation of spammers' social networks (Yang et al., 2012), extraction of sporting event summaries from Tweets (Nichols et al., 2012), and understanding the spread of information during times of social upheaval (Starbird and Palen, 2012). Twitter data has been used to explore patterns of discussion during emergency situations (Cassa et al., 2013), smoking behavior (Myslín et al., 2013), and many other health-related topics. Facebook has also been the subject of significant research interest, including studies of strengths of relationships (Xiang et al., 2010), relationships between social network use and well-being (Burke et al., 2010), and information diffusion (Bakshy et al., 2012) to name just a few. However, as for-profit businesses, Twitter and Facebook consider their data to be valuable, making only a subset available through APIs, with access to larger data sets possibly available for a fee (Finley, 2014). Twitter has also made limited access to their archives of historical content available to researchers through a data grant program (Kirkorian, 2014). Largely as a result of restrictions on data availability, this research is often conducted by researchers employed by the social networking sites being studied (Xiang et al., 2010; Burke et al., 2010; Bakshy et al., 2012).

Bulk datasets often make good data sources for studies of interaction patterns. Studies of Wikipedia trends have relied on bulk data downloads providing snapshots of site content at specific points in time (Viégas et al., 2007b)—such datasets can be invaluable when available, but the volume of content can also be daunting. Sampling of a smaller subset, either randomly, by time, or by content, can be an appropriate means of identifying a more manageable dataset. The Enron corpus, a database of several hundred thousand email messages from the failed energy company, provides an uncommon view into the electronic communications in a large company. This dataset has been analyzed in dozens of studies, addressing questions such as the identification of words and phrases used to indicate power relations in the corporate structure (Gilbert, 2012).

As with social network data, search engine research is perhaps most easily conducted by scientists working in the research labs of prominent search engine firms like Google (Ginsberg et al., 2009) and Microsoft (Huang et al., 2011, 2012; White and Horvitz, 2009; White, 2013; White et al., 2013; White and Hassan, 2014). See the "Google Flu" Sidebar for a discussion of the promises and challenges of log analysis, as illustrated by the high profile case of Google's Flu prediction analysis.

GOOGLE FLU

The history of Google's flu trend analysis tools (https://www.google.org/flutrends/about/) illustrates some of the potential value—and some of the pitfalls—in examining search data. Google's team analyzed a large corpus of search queries combined with geographical information identifying the location from which each query was issued. Noting a strong correlation between flu-related queries and clinicians' visits potentially related to flu, they were able to accurately predict which regions in the United States were experiencing flu outbreaks (Ginsberg et al., 2009). The excitement generated by these results was soon tempered by further experience demonstrating the trickiness of relating web search activity to online reality. A 2011 investigation of the performance of Google Flu Trends during the 2009 H1N1 influenza pandemic found that search behavior changed during the pandemic, as users searched for terms for influenza and related complications (Cook et al., 2011), and the estimates for the 2013 flu season varied radically from those issued by the Centers for Disease Control (Butler, 2013). A 2014 commentary reviewed related results and suggested that search data might be most useful when combined with other existing data sources (Lazer et al., 2014). This commentary also raised an important concern relevant to other studies of web search trends: as search engines are based on proprietary algorithms subject to regular revision, results may not be reliable or replicable (Lazer et al., 2014). Unsurprisingly, the exploration of twitter data for tracking flu epidemics has also been an area of active research (Allen et al., 2016; Santillana et al., 2015).

Despite concerns regarding the validity of predictions generated by Google Flu Trends, search logs continue to be a rich source of data for researchers interested in studying the implications of health-related terms. Some of this work attempts to validate Flu Trends, using other relevant indicators, such as flu-related visits to emergency departments (Klembczyk et al., 2016) as comparison points. A South Korean effort used social media (Twitter and blog) efforts to identify potential starting points in a subsequent examination of search terms for flu-related concepts (Woo et al., 2016), providing an example of the utility of combining multiple sources of online behavior data. Other efforts include flu tracking using only Twitter data (Allen et al., 2016; Santillana et al., 2015), and the use of search logs to identify possible adverse interactions between two drugs (White et al., 2013), to study the increasing severity of concern when searching for medical content (known as “Cyberchondria”) (White and Horvitz, 2009), or to identify symptoms that might be early indicators of cancers (Paparrizos et al., 2016). Related studies have used search data to explore biases in the search for health-related information (White, 2013; White and Hassan, 2014).

If your data source is either inaccessible due to business concerns, lack of an open API, or unacceptable costs, you might consider reframing your study to match what can be accomplished within your means. Substituting smaller scale studies or qualitative research for broad examinations into usage patterns might be one approach. One study used a set of interviews with Facebook users to understand how the content, layout, and functionality of the site influenced communication of health information (Menefee et al., 2016). Although smaller qualitative studies lack the broad appeal of the analysis of millions of posts, they might be more economical to complete.

If you are lucky enough to get your hands on a large dataset relevant to your interests, you might use a variety of techniques, depending on your interests and goals. Be prepared to spend some time on data cleaning and extraction, potentially taking textual representations of tweets, posts, or other data and formatting them in a normalized pattern suitable for querying or text searching (Baeza-Yates and Riberio-Neto, 2011). Once the data is ready for analysis, you may use any of a range of techniques. Possibilities include natural-language processing approaches that try to extract key concepts and relationships from free text (Hedegaard and Simonsen, 2013), and information retrieval techniques (Baeza-Yates and Riberio-Neto, 2011) to model similarities between documents and common concepts and terms. Other approaches have used descriptive statistics tracking types of activities and relationships (Kittur and Kraut, 2008), relative frequencies of different types of events (White et al., 2013), and any number of other techniques as appropriate. For social media analysis, you might build networks indicating relationships between individuals, topics, and other items of interest. Graph algorithms might be used to find network members who are "hubs"—outliers in terms of number of connections or presence on important paths (Scott, 2013). The Social Media Research Foundation (http://www.smrfoundation.org) has developed a tool known as NodeXL, which supports the development of networks, calculation of centrality measures, and visualization, all through spreadsheet data (Bonsignore et al., 2009; Hansen and Shneiderman, 2010).

In a refrain that should be familiar to readers who have made it this far, any of these data sources can be augmented by appropriate analysis with related data collected through different modalities. Examples include the use of surveys to understand user practices and beliefs with regard to searches for health information (White, 2013) and the use of instrumented web pages (Chapter 12) (Huang et al., 2012) or eye tracking (Chapter 13) (Huang et al., 2011) to capture fine-grain data correlated with search engine interactions. Approaches like these also open search engine interaction research to those who are not directly working with the relevant companies, as logging toolkits and eye-tracking experiments might be conducted in usability labs lacking access to large volumes of search interaction logs.

14.2.4 ONLINE RESEARCH DESIGN CHALLENGES

14.2.4.1 Appropriate topics for online research

Although it may seem somewhat obvious to note that online research will involve working with participants who are online, this helps point us toward the insight that online HCI research may be most appropriate for studies about the tools that people use online and the uses that they make of those tools. Participants in online studies will probably be working with web browsers, chat tools, and related online software as they read instructions, provide informed consent, perform tasks, and otherwise complete your experimental protocol. Research that works within this realm may be most successful.

Specifically, studies involving web applications or online tools are particularly well suited for online research. If you are running the web site on your own servers, web logs (Chapter 12) can provide useful feedback regarding timing, tasks, and errors. Conversely, studies of other application software, mobile devices, or novel interaction devices may be harder to do online: data collection is likely to be more difficult, incompatibilities between software versions may pop up, etc.

That is not to say that online studies of web site designs are easy. Good design practice certainly calls for cross-platform testing, but there is no guarantee that you will not run into versioning and compatibility problems, even with seemingly straightforward web pages.

14.2.4.2 Recruiting

By opening your research up to the Internet, you provide yourself with access to a much larger pool of participants. Recruiting can be easier, as emails to appropriate lists and postings on various web sites can go a long way toward identifying potential subjects. As online research generally involves the use of a web site or other online software, participants do not need to be local. Self-driven web site or study tools allow participants to complete tasks at their leisure, eliminating the need for scheduling.

Just as the use of undergraduates as study participants introduces a bias that may not be appropriate for some studies, online recruitment limits your subject pool to a particular segment of the larger population: Internet users who are interested enough to participate. This may mean that you might not attract relatively inexperienced individuals or participants who limit their time online to relatively focused activities. Whether or not this poses a problem depends on the specifics of the study in question.

In some cases, online research can give you access to pools of participants that otherwise would have been unavailable. This is particularly true for people with disabilities, who may find traveling to a researcher lab to be logistically unfeasible (Petrie et al., 2006), and domain experts, who may be hard to find in sufficient numbers in some locales (Brush et al., 2004). See Chapter 16 for more details on HCI research involving people with disabilities. Collaborative research involving distant partners can also be substantially aided by online tools for communicating and gathering data.

One important difference between online and in-person research is the potentially complete anonymity of participants in online studies. When you meet a participant face-to-face, you can usually make a pretty good guess about their age, gender, and other demographic characteristics. The lack of face-to-face contact with online participants makes verification of such details harder—you have no way of verifying that your participants are male or female, old or young. This presents some recruiting challenges, particularly if your research requires participants who meet certain demographic constraints such as age or gender. If your only contact is via email or other electronic means, you may not be able to verify that the person with whom you are communicating is who he or she is claiming to be. Online studies that do not require the participants to reveal their true identity (relying instead on email addresses or screen names) are highly vulnerable to deception. Certain incentives, such as offering to enter participants in a draw for a desirable prize, might compound this problem. For example, a survey aimed at a specific demographic group might draw multiple responses from one individual, who might use multiple email addresses to appear as if inquiries were coming from different people. Possible approaches for avoiding such problems include eliminating incentives; requiring proof of demographic status (age, gender, disability, etc.) for participation; and initial phone or in-person contact in order to provide some verification of identity. Since payment or other delivery of incentives often requires knowing a participant's name and address, verification of identity is often not an added burden.

14.2.4.3 Study design

Surveys (Lazar and Preece, 1999) (Chapter 5), usability evaluations (Brush et al., 2004; Petrie et al., 2006), and ethnographic studies of support groups (Maloney-Krichmar and Preece, 2005) have all been successfully completed online. Examples of online usability studies have shown that both synchronous studies with domain experts (Brush et al., 2004) and asynchronous studies with users with disabilities (Petrie et al., 2006) have yielded results comparable to those that were found in traditional usability studies. Perhaps due to difficulties in sampling and controls, online empirical studies of task performance are less common. One study of the influence of informal “sketch-like” interfaces on drawing behavior used an online study as a means of confirming the results of a smaller, traditional study. Results from the 221 subjects in the online study were highly consistent with the results from the 18 subjects in the traditional, controlled study in the lab. The agreement between the two sets of results provides a more convincing argument than the lab study on its own (Meyer and Bederson, 1998).

Opinions differ on the appropriateness of online research for different types of data collection. The lack of controls on the participant population might be seen as a difficulty for some controlled, empirical studies. Others have argued that as online research does not allow for detailed user observation, it is more appropriate for quantitative approaches (Petrie et al., 2006). In the absence of any clear guidelines, it is certainly appropriate to design studies carefully and to clearly describe and document the reasoning behind any designs that are adopted. When possible, hybrid approaches involving both in-person and online research may provide additional data and avoid some of the downsides associated with each approach.

Online studies involving surveys, self-selected visits to web sites, crowdsourcing, or other approaches that do not require synchronous interactions with researchers might be subject to frequent dropouts, as users decide to start a task and then stop half-way through. Study designs should anticipate such dropouts and consider how they might be reported. If you are looking at task completion success rate, it is probably appropriate to include all participants who started the task. If you are looking at task completion times, you might want to focus only on those who completed the tasks. Providing numbers for those who started tasks, those who completed tasks, and indicating which groups were considered for which analyses is probably most appropriate.

14.2.4.4 Ethical concerns

Although the usual guidelines regarding protection of participants apply to online research, numerous confounding factors can create some interesting and challenging dilemmas.

Studies of online communities must consider questions of privacy and online consent. What is the expectation of privacy when participants in an online forum post messages publicly? Are such messages fair game for researchers? Is informed consent required before messages can be used? What if the site is only accessible to users who register and login? These questions have generated debate, discussions, and some guidelines (Bruckman, 2002; Frankel and Siang, 1999), but specific issues vary from case to case. Researchers are urged to be particularly careful when exploring communities describing sensitive topics such as health. The trust needed for participants to share stories of challenging personal times such as illnesses may lead some users to forget that they are effectively participating in a public forum where materials may be read by many individuals. Lurking in such communities or posing as a member may not be seen as appropriate behavior. Before doing so, you might consider talking to the organization or individuals responsible for the site and introducing your study to the group. Creating communities specifically for research purposes can be a successful—if not always practical—alternative (Bruckman, 2002).

Informed consent and debriefing for online studies can also be tricky. Providing important information for either of these tasks via online text may not be sufficient. In-person studies provide the possibility of direct feedback: experimenters know if participants have any questions or if there is any postexperiment distress. These factors are much harder to gauge online (Azar, 2000). Although one study indicated that comprehension of informed consent forms online may be comparable to comprehension of forms on paper, poor recall in both cases illustrates the general challenge of constructing effective consent forms (Varnhagen et al., 2005). These issues may be even thornier for studies conducted retrospectively, through API access to posted data or other methods allowed under web site terms of service. Although such studies are not inappropriate, and may not require consent, it is still best to tread carefully. When possible, provide clear and easily understandable descriptions of research goals and implications. In any case, these studies should not be undertaken without careful attention to appropriate rules for protection of human research participants. Detailed discussions of human subject protections can be found in Chapter 15.

The considerable challenges and headache associated with deceptive online research provide a strong argument against this sort of approach. If you find yourself tempted to try this sort of study, consider a lab-based study instead. You may still use deception in this case but the use of prior informed consent can help you avoid many difficult questions.

As with any HCI research, online research can be particularly challenging if there is potential harm involved or when dealing with special cases, such as research involving children. Technical measures such as encryption of transmitted data may be useful for privacy protection and for verifying parental consent in the case of minors (Kraut et al., 2004). Laws such as the Children's Online Privacy Protection Act of 1998 (COPPA) in the United States may limit the amount of information that can be collected from minors. Researchers working in these areas should construct study materials carefully; consult with appropriate authorities responsible for human research participant protection (known as Institutional Review Boards in the United States—see Chapter 15) and external experts to review proposed procedures; and use traditional studies as opposed to online studies when appropriate (Kraut et al., 2004).

14.3 HUMAN COMPUTATION

14.3.1 INTRODUCTION TO HUMAN COMPUTATION

What can people do more effectively than computers? Despite the frustrations associated with seemingly endless bugs and glitches, most people who use computers frequently would probably agree that computers do many jobs more quickly and more accurately than humans (if you ever talk to someone who disagrees, see what happens if you ask them to give up their smartphone or laptop). However, there are some areas where humans continue—at least for the time being—to outperform computers. Tasks requiring detailed interpretation of complex inputs are a prime example. Despite recent improvements in computer vision, natural-language processing, and other fields of artificial intelligence, software systems often struggle to identify objects in digital images or to interpret written text, even when such tasks are straightforward for many humans.

Given these differing—and often complementary—strengths of both humans and computers, many observers have argued for the use of computers to augment human cognition (Shneiderman, 2002). This line of inquiry dates back to the prehistory of HCI, in speculative designs such as Vannevar Bush's Memex (Bush, 1945) and Douglas Englebart's work on augmenting human intellect (Engelbart, 1962), which led to the famous 1968 demos of the first computer mouse, early graphical user interface, and word processor.

Human Computation takes the opposite approach. Given a task that might be hard for a computer but relatively easy for a human, a human computation strategy might ask multiple humans to complete small pieces of that task. For example, consider a computer vision algorithm for identifying numerals in digital photographs. A machine learning tool for such a task might be challenged by the range of sizes, fonts, and colors of numerals found on building, signs, and elsewhere in images, even when a human could read those numbers very easily. A human computation task might ask multiple participants to interpret a large set of images, thus providing a large collection of labeled images. Resulting labels might be used to train improved machine learning for classifying similar images, or to develop a search tool for identifying images matching specified descriptions. These tasks that require human—as opposed to computer—cognition are often referred to as *Human Intelligence Tasks*. When such tasks are explicitly organized with the goal of efficiently finding an accurate solution for a computational problem, the resulting system might be called a *human computation system* (Law and Ahn, 2011). See the "CAPTCHA and reCAPTCHA" sidebars for the story of the most familiar human computation tasks.

CAPTCHA AND reCAPTCHA

CAPTCHA—the Completely Automated Public Turing test to tell Computers and Humans Apart—is perhaps the most familiar example of human computation. The term CAPTCHA was developed by Luis von Ahn and colleagues, who proposed the use of a problem that is hard for computers but easy for humans as a web site security measure, suitable for distinguishing between human visitors to a site and automated scripts pretending to be humans (Ahn et al., 2003). The original task—deciphering letters in a word distorted so as to defeat computer vision programs—has since spawned numerous variations familiar to users of many web sites.

A closely related line of research explored related ideas, originally in the realm of image annotation. Annotations in the form of image labels are required to support image search, as computer vision tools may not be sufficiently powerful to identify image content matching terms of interest. However, these labels are not easy to come by, as they must be generated by humans who must interpret the images and provide descriptions. Noting these problems, Luis von Ahn and Laura Dabbish suggested a simple and intriguing solution: turn it into a game. The ESP game presents two players with an image, asking them both to provide a label describing the image. The players are challenged to come up with an agreed-upon description, getting points for each agreement, with large bonuses for surpassing a certain goal in a given time period. The need for agreement creates the challenge that makes the game enjoyable, while increasing the quality of the labels, as two participants are unlikely to agree upon an inaccurate description. Additional labels can be generated for each image through the use of "taboo" words: once a first pair of partners labels an image, subsequent partners will be asked to find a label without using any of the previously used words (Ahn and Dabbish, 2004). The ESP game introduced the notion of "Games with a purpose"—tools that hide useful work under the guise of a challenging and enjoyable game (Ahn and Dabbish, 2008). Just as Tom Sawyer turned the work of painting a fence from a chore into a pleasure, games with a purpose turn image labeling and other tedious tasks into a bit of fun.

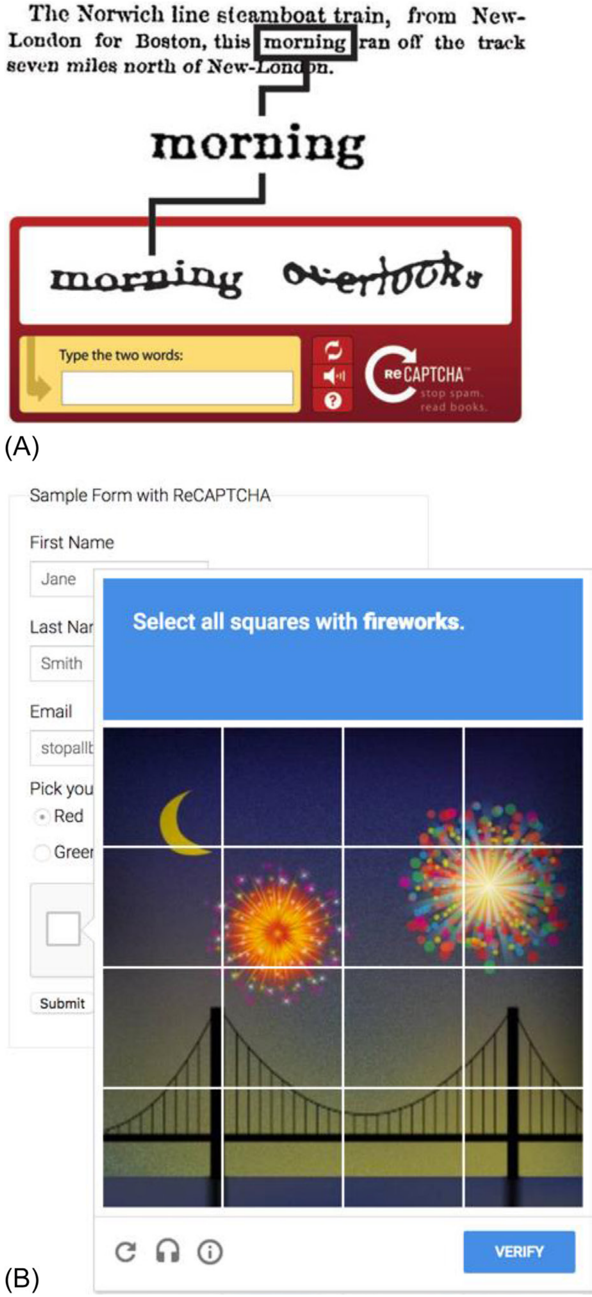

FIGURE 14.1

reCAPTCHA: (A) The original reCAPTCHA asked users to type in words that could not be recognized via optical-character recognition. A sample of text that might have been recognized is shown, along with a depiction of how the text might be distorted before being presented to the user. (B) More recent reCAPTCHA tasks involve image classification, such as choosing images from a set that match a specified criteria—in this case, images containing fireworks.

(A) From von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M., 2008. ReCAPTCHA: humanbased character recognition via web security measures. Science 321, 1465–1468; (B) From https://www.google.com/recaptcha/api2/demo (accessed April 8, 2017).

(Continued)

CAPTCHA AND reCAPTCHA—CONT'D

Subsequent work merged CAPTCHA's goal of using human intelligence tasks as security with the ESP games notion of using these tasks to accomplish useful work, leading to the reCAPTCHA tool (von Ahn et al., 2008). reCAPTCHA was designed to solve the problem of digitizing text that had proven challenging for optical-character recognition (OCR) systems. reCAPTCHA provides users with images including text that has proven difficult for computer vision systems to interpret. Specifically, the original reCAPTCHA asked users to decipher words that have each failed to be consistently recognized by two different OCR programs. Each time a reCAPTCHA is used, the user is asked to interpret images containing two words: one for which the interpretation is known, and another which has not yet been classified. If the user provides a correct answer for the known word, the answer for the other word is assumed to be correct. Each word is presented to multiple users, and words can be promoted to become known words if sufficient accurate human guesses are provided. All words are distorted in an attempt to defeat computer vision programs (Figure 14.1A) (von Ahn et al., 2008). reCAPTCHA has been used on many web sites to provide the security that motivated the design of the original CAPTCHA, primarily verification of user registration and login on web sites. reCAPTCHA was purchased by Google in 2009 (Zlatos, 2009), with subsequent evolution of the tool including variants for labeling images (Figure 14.1B) and predictive tools capable of identifying users as human based on interactions with the widget, without the need for image labeling (Shet, 2014).

reCAPTCHA's use of images highlights a key design challenge. The image-labeling tasks in the ESP game were purely entertainment on the part of the users. CAPTCHAs, on the other hand, are often used on sites that might be the sole route for users to access functionality needed for personal or professional purposes. As a result, accessibility becomes a key concern, as some users—particularly those with low vision or blindness—might struggle with some of the images used in tools like reCAPTCHA. This problem is magnified by the nature of the tools—by definition, the images used in reCAPTCHA are those that have been in some ways hard to process. reCAPTCHA has always had an audio option, which has generally asked users to type a sequence of spoken digits. Alternative CAPTCHA tests have been the subject of multiple research efforts (Sauer et al., 2010; Davidson et al., 2014).

Although reCAPTCHA is likely the most familiar human computation task, the notion of using games to motivate participation has been used in many different domains. Online games have been particularly successful in scientific fields, with the Fold.It game (http://fold.it) harnessing the power of multiple users to generate high-quality protein models (Khatib et al., 2011; Eiben et al., 2012) and bioinformatics games at http://www.genegames.org challenging users to complete tasks such as curating gene-disease associations (Good et al., 2012).

Despite the success of games with a purpose and related tools, not all tasks in need of human input are easily converted into small subtasks amenable to competition or collaboration between participants. Longer, more complex tasks may take more time to complete and require additional training or expertise. *Crowdsourcing studies*[1] use online platforms to collect data from participants over the web, usually through the use of web software designed to enroll participants, provide training, and complete relevant tasks.

Crowdsourced research studies can be (roughly) divided into two key groupings. Studies involving *systems based on crowdsourced data* explore applications of user-contributed data to develop novel solutions to challenging problems. Like CAPTCHA and other human computation tasks described earlier, these studies are all focused around some task(s) that humans can do better than computers. Examples include annotating research reports to identify discussions of potentially harmful drug-drug interactions (Hochheiser et al., 2016), extracting relationships between texts and tables in written reports (Kong et al., 2014); delivering crowd-based emotional support in online interventions for depression (Morris et al., 2015); translating text (Hu et al., 2014); prototyping user interface designs (Lasecki et al., 2015); and using real-time crowd interpretation of cell phone images to help blind people identify nearby objects (Bigham et al., 2010; Lasecki et al., 2014), to name just a few of many.

A second, crowdsourced model involves *crowdsourced HCI experiments*: web-based studies involving large numbers of participants in more or less traditional empirical evaluations of interfaces or visualizations. As the goal of these studies is to evaluate how humans use a tool to accomplish a task, they are not necessarily strictly human computation: some studies in this category may include tasks that might, in fact, be done by computers. However, other elements are similar, in that large numbers of people will be asked to complete tasks, through an online infrastructure supporting with recruitment, enrollment, and data collection. Examples of crowdsourced experiments have been used in studies evaluating visualization designs (Heer and Bostock, 2010; Abdul-Rahman et al., 2014; Micallef et al., 2012), mobile applications (Zhang et al., 2016), and even (via a creative proxy) haptic interfaces (Schneider et al., 2016).

14.3.2 CONDUCTING HUMAN COMPUTATION STUDIES

Using crowdsourcing services to inexpensively identify and enroll a large pool of study participants might appear to be a very appealing prospect. However, matters are (perhaps unsurprisingly) not quite that simple, as previous work has identified concerns that might impact the quality of the data collected. Consideration of these concerns, and of recommendations originating in these earlier studies, can help you design tasks and use task performance data to ensure that your experiments generate the high-quality data that you need to move your research.

[1] Not to be confused with crowdsourcing content, which refers to the process of combining the efforts of multiple authors and editors to write articles such as those found on Wikipedia.

To understand the challenges, we might compare crowdsourced studies to traditional studies. Familiar lab-based studies use advertisements and word of mouth to spread the word, often offering a small honorarium to encourage interest. Participants come to the lab, spend some amount of time—perhaps an hour or two—and are given payment upon completion of their participation. Although this approach often leads to maddening difficulties in recruiting sufficient numbers of participants, it offers several advantages. Perhaps most importantly, individuals who express interest in such studies can usually be depended upon to complete the studies appropriately and in good faith. Enticements of $20 or even $50 might be sufficient to encourage some people to participate in studies that do not interest them, but it is generally not worth the bother to participate without taking the study seriously. Although we have not evaluated the question empirically, our experience has been that most people who agree to join in lab studies do so honestly and with every intention of working with the researcher to meet the goals of the study.

Direct interaction with participants is a second, closely related, benefit. When someone sits down in your lab to participate in a study, you will be able to talk with them and to observe their work as they complete the tasks at hand. These interactions provide valuable "sanity check" information, allowing you to form impressions of each individual's task performance and motivations, and specifically to avoid participants who might not be taking your tasks seriously. You certainly do not want to rush to discard data from someone who is goofing off—including the data and raising the concern in a discussion would be much more appropriate—but having observed this behavior might help you understand results, particularly if you identify participants with bad behavior that might have led to unexpected results in your data.

There are many appealing aspects to the use of human computation in HCI research. A properly constructed human computation study can be constructed in software, deployed on a web site (often using dedicated software services, as discussed later), and advertised to large numbers of potential workers at reasonably low cost. Participant enrollment, completion of consent forms, administration of the study, and data collection can be largely automated, thus eliminating the need for tedious work that has afflicted many graduate and undergraduate student workers. Online human computation studies can also enroll many more participants than comparable traditional studies, providing greater statistical power. The user base may be large and diverse, involving a broader range of education levels, ethnicities, and backgrounds than you would likely get in a lab (Kittur and Kraut, 2008).

Of course, the reality is somewhat more complicated. As with any other type of HCI study, human computation experiments require careful selection of participants and tasks. You will also need an appropriate software infrastructure, capable of handling all of the enrollment and screening processes conducted to enroll participants; and the presentation of tasks and collection of data necessary for the study itself. Human computation studies must be carefully designed to ensure high-quality responses: although tasks involving intrinsic motivation such as entertainment, intellectual curiosity, or accessing a desired resource might motivate participants to perform honestly and accurately, users who are paid for their answers (as is often the case) might be in it for the money. The nature of the small tasks might encourage users to emphasize speed over accuracy, rushing through tasks to collect as much payment as possible, without any regard for the quality of the answers provided (Kittur and Kraut, 2008).

14.3.2.1 Software infrastructure

Human computation studies need not have extensive or complex software infrastructure. Studies can easily be run through homegrown or customized web applications, together with logging software capable of tracking the details and time of any given interaction. One productive approach for such tools might be to build a database-driven web application capable of storing appropriate demographic background information associated with each participant, along with details of each action and task completed. You might even add an administrative component capable of managing and enrolling prospective participants. These homegrown applications are generally not terribly difficult to construct, particularly if you have a web-based implementation of the key tasks under consideration, or were planning on building one anyway. For some tasks—particularly those involving collection of fine-grained detail or requiring complex interactions—the freedom associated with constructing your own application may be necessary to get the job done.

Commercial crowdsourcing services provide an attractive alternative to homegrown software. These commercial offerings provide platforms for creating tasks, including providing training materials, presenting task components, and collecting task results, with tools designed to minimize—if not eliminate—the need to do any programming. Perhaps even more importantly, they also offer access to registered workers who have expressed interest in completing small tasks in exchange for *micropayments*. This infrastructure significantly simplifies recruitment of participants—once you publish your tasks registered users can find them on the site and get to work. These tools can facilitate enrolling users, managing payments, and even prescreening users to verify eligibility in terms of demographic requirements (gender, age, etc.) or background knowledge (Paolacci et al., 2010), thus eliminating many of the headaches of study design. Although Amazon's Mechanical Turk (http://www.mturk.com) is by far the crowdsourcing tool most used in published human-computer interaction studies, other systems such as CrowdFlower also appear in the literature (Kucherbaev et al., 2016).

Although details obviously differ across platforms, construction of studies is generally straightforward. Tasks and instructions can be created via tools provided by the sites, with custom HTML and JavaScript programming as needed, particularly for more complex tasks. Some platforms also allow tasks that load contents of external web sites (McInnis and Leshed, 2016), providing more control to task designers. Software development APIs often provide additional flexibility, at the cost of some amount of programming (Amazon, 2016; CrowdFlower, 2016).

A number of research efforts have extended the Mechanical Turk software toolkits to better support crowdsourced studies of web interface usability (Nebeling et al., 2012, 2013); to enable synchronous and longitudinal studies (Mao et al., 2012); to use Turk workers to plan the contents of tasks to be completed by subsequent workers (Anand et al., 2011); and to simplify the construction of tasks (Greg et al., 2010), potentially including components for predicting and evaluating confidence level and costs (Barowy et al., 2016).

The infrastructure provided by Mechanical Turk and similar crowdsourcing platforms provides many advantages over “roll-your-own” designs. As any experienced HCI researcher knows well, the challenges of recruiting, enrolling, and consenting participants can consume substantial amounts of time. Even if you are able to build your own web application to do the trick, you might find that leveraging these platforms—particularly with one of the add-on libraries—might simplify your life considerably. These advantages aside, commercial crowdsourcing tools have potential downsides. Financially, payment for microtasks might be more expensive than the gifts or small payments traditionally made to study participants. Be sure to estimate your costs before you embark on a study. Technical challenges may arise—integration of complex, preexisting web applications with APIs provided by the crowd worker platforms might be a complex task. Consider a preliminary study to prove the concept and test the tasks thoroughly before starting a study. If the commercial platform does not work out, you might want to fall back on homegrown tools. In any case, you will still have to deal with the selection of which tasks you want users to complete and how you might design the tasks to ensure high-quality responses.

14.3.2.2 Tasks and study design

Law and von Ahn present a framework for developing appropriate tasks for human computation studies (Law and Ahn, 2011). Tasks can be seen as containing three main elements: introductory description, clear definitions of success criteria, and incentives (financial for Mechanical Turk and other systems, entertainment for games, access to services for CAPTCHA). Each task will involve multiple design decisions, including which *information* is presented to encourage completion of tasks without bias; tradeoffs in *granularity* between the value of the result and the time required to complete; whether tasks are completed individually or collaboratively; which incentives are offered, and how quality is ensured (Law and Ahn, 2011). For an in-depth discussion of these and related issues, Law and von Ahn's in-depth discussion (Law and Ahn, 2011) is highly recommended. An alternative model is presented by Alexander Quinn and Benjamin Bederson, who developed a multidimensional classification taxonomy. The Quinn-Bederson model describes human computation systems in terms of motivations for participation, quality control measures, techniques for aggregating responses, required human skills, orders and workflows for processing tasks, and the cardinality of tasks to requests (how many users are mapped to each task) (Quinn and Bederson, 2011).

Concerns over quality control have led to a variety of approaches in task design to attempt to ensure high-quality results from crowdsourcing studies (Table 14.1).

Table 14.1 Quality Control Measures for Crowdsourcing Studies

Strategy	**Proposed Approach**
Question design	Include questions with known answers (Kittur et al., 2008)
	Make accurate answers easy to provide (Kittur et al., 2008)
Study design	Develop predictive models based on question types to determine how many responses are needed to ensure high-quality answers for each question type (Barowy et al., 2016)
	Use micro-diversions or other distracters to offset declines in response quality as users get bored or tired (Dai et al., 2015)
Task performance data analysis	Look for patterns indicating answers that might have been faked or rushed, including repeated free text or questions answered too quickly (Kittur et al., 2008)
	Use task completion metadata to develop predictive models of individual workers (Ipeirotis et al., 2010) and tasks (Rzeszotarski and Kittur, 2011; Zhu et al., 2012)

Anniket Kittur, Ed Chi, and Bongwun Suh (Kittur et al., 2008) made three suggestions for designing high-quality crowdsourcing tasks. (1) Each task should include questions with known answers that can be easily checked. Asking participants to count the number of images in the page, or to answer a simple question based on the text in the page, can help determine if they are answering seriously or simply rushing through. (2) Accurate answers should be no harder to provide than rushed, inaccurate answers. For example, a task asking users to summarize a site might be easily subverted by short one-word answers, but an explicit requirement that users provide a certain number of keywords to describe content might be easier to fill out accurately. (3) Look for other ways to find low-quality answers, such as by identifying tasks that are completed too quickly or have answers repeated across multiple tasks (Kittur et al., 2008). Having multiple users complete each task and using agreement on results as a measure of quality—just as described earlier for CAPTCHA—is another possibility, but redundancy can be expensive (Ipeirotis et al., 2010). Alternatively, models of the complexity of different response types (checkboxes, radio boxes, free text) can be used to predict the number of responses needed to arrive at high-quality levels with high confidence (Barowy et al., 2016). "Micro-diversions"—games or other entertaining distractions designed to disrupt the monotony of performing multiple repeated tasks over long periods of time—might also help improve response quality (Dai et al., 2015).

Other studies have used task completion metadata to develop predictive models suitable for identifying invalid answers. Noting that Mechanical Turk collects detailed data on each task, including measures of start and end time, Zhu and colleagues built predictive models based on initial estimates of task performance and data from actual tasks. They then used these models to classify subsequent responses as either valid or invalid (Zhu et al., 2012). Other efforts have explored building models of individual workers (Ipeirotis et al., 2010) and using JavaScript web-page instrumenting techniques (Chapter 12) to collect mouse and keyboard usage data sufficient for building "task fingerprints" capable of predicting performance (Rzeszotarski and Kittur, 2011).

Successful design of a crowdsourced study does not end with the design of individual tasks. Although some studies—particularly studies involving online evaluation of user interface designs—may be based on large numbers of workers completing very similar tasks, more complex control structures have been used in crowdsourcing studies to decompose large problems, to introduce feedback—whereby responses to some questions will influence the content of subsequent question, or to influence workflows. Edith Law and Luis von Ahn provide a summary of different workflow strategies in their in-depth review of human computation (Law and Ahn, 2011).

14.3.2.3 Pros and cons of crowdsourced studies

Easy to create, potentially inexpensive, and backed by services that simplify recruitment and enrollment of participants, crowdsourced studies can be very appealing. Other potential advantages include potentially decreased bias and increased validity, as participants who do not interact directly with researchers or even know that they are participating in an experiment might be less susceptible to implicit or explicit pressures (Paolacci et al., 2010). Although the use of services like Mechanical Turk does remove some knowledge about participants (Kittur et al., 2008), some have argued that Turk users may be demographically similar to broader populations (Paolacci et al., 2010). Technical questions might influence the validity of task completion times from crowdsourced experiments, as network delays might impact task completion times (see Chapter 12). Finally, the lack of direct interaction with participants eliminates the possibility of gaining any insight from direct observation of task completion. Pairing studies—as discussed earlier—provides one possible means of avoiding this lack of feedback. A small lab study might give you the insight associated with direct interaction with users, while a companion human computation study will help you enroll larger numbers of participants.

Before jumping into studies using systems like Mechanical Turk, you should take care to ensure that your software components are implemented and tested correctly, and that you understand the social dynamics of the workers. Online forums for mechanical Turk users, including Turkopticon (https://turkopticon.ucsd.edu) (Irani and Silberman, 2013) and Turker Nation (http://turkernation.com), provide workers with the opportunity to discuss interesting tasks, problems with task requestors, and other topics of interest to workers trying to earn money through Mechanical Turk. These groups can provide valuable resources and feedback to researchers using human computation in their work. Brian McInnis and Gilly Leshed described how interactions with these groups proved particularly useful when software errors prevented tasks from working correctly, and workers from being paid. Interactions with the participant community helped resolve the issues and provide fair payment, thus avoiding the unfortunate outcome of a failed experiment leading to ill will (McInnis and Leshed, 2016). It may not be possible to identify all problems in advance, but working with the community of users to build trust and promote fairness may be an important strategy for successful human computation studies.

14.3.3 FUTURE OF HUMAN COMPUTATION

Human computation has many promising applications. A 2014 workshop of the Computing Community Consortium of the Computing Research Association outlined numerous possibilities for the use of human computation to meet pressing social needs, including social support for people in need; combining training with problem solving to improve the process of interpreting radiology images; to collect river-level information and serve as an early warning for possible floods; and others (Michelucci et al., 2015). As our engagement with our devices continues to occupy much of our time and attention, attempts to channel this fascination in socially meaningful ways are likely to continue to be a growing part of the landscape.

HCI research efforts have explored possible extensions to crowd source models, designed to increase the utility of crowdsourced work. Possibilities include changing task structures to include "handoffs" between workers, thus possibly increasing the quality of the resulting work (Embiricos et al., 2014); using algorithmic approaches to plan task workflow (Weld, 2015); exploring the impact of task ordering on speed and mental demand during the completion of a sequence of small tasks (Cai et al., 2016); and using new models to encourage participation, including leveraging participant curiosity (Law et al., 2016), providing entertaining "micro-diversions" to improve productivity of workers conducting many tasks (Dai et al., 2015), or using "twitch" microtasks capable of being completed very quickly to lower barriers to involvement (Vaish et al., 2014). Other efforts have explored paying crowd workers to be ready to respond quickly, thus enabling real-time crowdsourcing (Bernstein et al., 2011), applying algorithmic approaches to identify when tasks should be reassigned because original workers have abandoned them (Kucherbaev et al., 2016), and using models of increased error tolerance to increase the rate at which large tasks can be completed (Krishna et al., 2016).

Another promising line of research asks a slightly different question—"how can crowdsourced workers help with familiar, knowledge-intensive tasks?" As complex tasks, writing papers, drawing figures and diagrams, and analyzing budgets require significant cognitive effort and attention to detail, crowdsourced workers might help writers, designers, and analysts with on-demand suggestions for improving the quality of their work. These possibilities drove the development of Soylent, a set of tools for using human computation to improve the writing process. Developed as extensions to Microsoft Word, Soylent provides writers with the ability to request human computation assistance in shortening texts, grammar and spell-checking, and other tasks not easily accomplished via existing word processing tools (Bernstein et al., 2015). Although the possibilities of using human computation assistance to assist with these familiar tasks may be intriguing, further work will likely be needed to understand when, where, and for whom such models are appropriate. For more news on developments in Human Computation, see the web site of the Human Computation Institute (http://humancomputation.org).

14.4 SENSORS AND UBIQUITOUS COMPUTING

Taking advantage of advances in miniaturization of components, reduced power requirements, and advances in abilities to sense and distribute information without physical connections through electrical or data networks, engineers have developed approaches that revolutionize our ability to collect data. Originally discussed in the context of the RFID tags that can be used to sense uniquely identified objects through radio frequencies, the "Internet of Things" (Ashton, 2009) has become a familiar means of describing a landscape where data collection, sensing, and computing are all around us, and often invisibly hidden in unobtrusive devices. Although the smartphones and fitness monitors described in Chapter 13 are perhaps the most familiar, they are only the beginning. Sensors embedded in clothes or eyeglasses have arrived in commercial products, and Internet-connected thermostats, security alarms, and security cameras help concerned homeowners keep an eye on things while they are away. Low-cost development platforms including Arduino and Raspberry PI provide tinkerers with the tools to design their own ubiquitous data collection tools.

Broadly speaking, HCI researchers engage in two types of research with these sensors. From a system-building point of view, needs assessment through qualitative research is generally needed to understand what should be built and how it should work. Once systems are deployed, analysis of interaction data (using techniques from Chapters 12 and 13) will generally be combined with ethnography (Chapter 9), case studies (Chapter 7), and other qualitative approaches (Chapter 11) to understand how the tools worked in practice. A discussion of some example systems and their techniques and methods will help us appreciate some of the challenges and how they have been addressed. Table 14.2 provides an overview summary of types of sensor/ubiquitous research, research methods, and challenges.

Table 14.2 Overview of Study Types, Research Methods, and Challenges of HCI Research Involving Sensors and Ubiquitous Computing

Study Types	Methods	Challenges
• Alternative input • Sensors and monitoring • Mobile devices • Wearables	• Diaries • Interviews • Field studies • Usage log analysis • Ethnography • Participatory design	• Data transfer • Data storage • Data processing • Engineering and configuration of sensors and networking

14.4.1 HISTORY AND EXAMPLES

The origins of ubiquitous computing are generally traced back to Marc Weiser's 1991 article "The Computer for the 21st century" (Weiser, 1991). Writing as chief technologist of the Xerox Palo Alto Research Labs—home of pioneering work in early graphical user interfaces—Weiser imagined a future working environment with active badges, computational devices, and pads and tabs linked by networks of wired and wireless transceivers. Perhaps more importantly, Weiser urged the notion of "embodied virtuality"—taking computing from the workstations of the time and embedding it in numerous, ubiquitous devices. Coming at a time when sensors and wireless communication were beginning to be sufficiently small, powerful, and inexpensive to enable this computing infrastructure (Weiser, 1991). Weiser's vision inspired researchers to create the field of Ubiquitous computing, led by the flagship UbiComp conference since 2001. A 2011 study of research and co-citation in ubiquitous computing identified almost 6000 papers published between 1995 and 2009 (Zhao and Wang, 2011)—a number that has undoubtedly grown steadily since (Figure 14.2). Although a full review of these varied research areas is beyond the scope of this section, description of some selected research efforts will identify recurrent trends and related research methods.

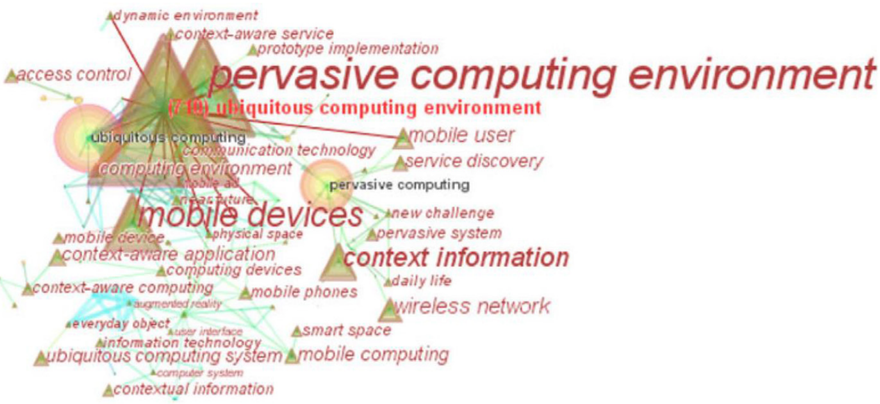

FIGURE 14.2

A map of keywords from titles and abstracts *(triangles)* and other noun phrases *(circles)* found in a collection of almost 6000 ubiquitous computing research articles published between 1995 and 2009.

From Zhao, R., Wang, J., 2011. Visualizing the research on pervasive and ubiquitous computing. Scientometrics 86 (3), 593–612.

Alternative input devices were an early focus of ubiquitous computing research. Noting the short comings of mouse and keyboard input, these projects explored various physical devices, complete with associated electronics, that might be used to control software applications. The *Tangible Bits* effort described several possible mechanisms including an arm-mounted LED screen, a horizontal back-projected computing "surface," and physical icons—known as "phicons"—that might be used to represent and manipulate items in the display (Ishii and Ullmer, 1997).

Sensors and monitoring tools have been developed to use computing to address challenges associated with everyday lives. Noting the difficulties that many families face in dealing with aging relatives who may want to remain in their own homes despite increased frailty and need for support, Elizabeth Mynatt and colleagues developed the digital family portrait, which would use sensors distributed throughout the house of an older person to collect data associated with successful completion of activities of daily life. These readings would be collected electronically and distributed via the Internet to a computerized picture frame in the home of the younger family member. A border on the screen would display icons that might vary in size and intensity to indicate recent activity levels and trends, assuring the younger caretaker that their older relative was safely up and about, while maintaining privacy for the older person (Figure 14.3) (Mynatt et al., 2001; Rowan and Mynatt, 2005). Similar concerns about the well-being of older people have led to designs combining motion sensing data from smartphones with Microsoft Kinect motion sensing (see Chapter 13) and social media feeds to identify activities in the home and related variations in mood (Ghose et al., 2013). Other studies have been more purely formative, such as a contextual interview study that asked older adults about objects of

FIGURE 14.3

A digital family portrait, with a picture of an older relative surrounded by butterfly icons scaled to indicate relative levels of activity. Levels for the current day are represented with new icons added hourly, while the previous 27 days are summarized with one icon per day.

From Rowan, J., Mynatt, E.D., 2005. Digital family portrait field trial: support for aging in place. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Portland, OR, pp. 521–530.

importance, in the hopes of developing technologies that might promote social interaction. Results of these interviews indicated that many valued devices were routinely used in social interaction, even though they were not specifically communication devices (Vaisutis et al., 2014).

Other efforts have explored the possibility of widespread data collection through mobile devices. The "Smart Citizen Kit" provides Arduino-based hardware, including sensors, battery, and WiFi, suitable for collecting environmental data that can be shared, aggregated, and used for community planning (Diez and Posada, 2013). Noting that data from these sensors often provide data that is hard to interpret, the Physikit project augmented the smart citizen kit with physical ambient displays (inspired, in part by Tangible Bits (Ishii and Ullmer, 1997)) using light, movement, buzzing, or air motion to display data trends (Houben et al., 2016).

From a research (but not commercial) perspective, it is hard to talk about ubiquitous computing without mentioning wearables such as Google Glass. Growing out of a long line of research on wearable computing, Glass was the first broadly available head-mounted display for everyday use. Although not commercially successful, Glass had an impact on the research world. Researchers have applied Glass to challenges such as object recognition (Chen et al., 2015) and gaze recognition (Kangas et al., 2014), with applications including assistive help for people with Parkinson's disease (McNaney et al., 2014), supporting hands-free computer use in laboratory settings (Hu et al., 2015), physics education (Weppner et al., 2014a), logging of activity (Weppner et al., 2014b), and encouraging exercise (Nguyen et al., 2014; Sörös et al., 2013). One prototype system combined Google Glass with an EEG headset (Chapter 13) to support home control, allowing users to combine gaze at an object along with a thought about a desired outcome to complete tasks such as adjusting a thermostat (Simoens et al., 2014). Google Glass may not have been a commercial success, but these and related efforts suggest that some future product might find its niche.

14.4.2 UBIQUITOUS COMPUTING RESEARCH METHODS

Ubiquitous computing research might be technologically innovative, but the research methods involved are often somewhat familiar, using methods described throughout this book. Given the importance of the embedded nature of these tools, qualitative case studies and other examinations of the use of the tools in context may be more prominent than empirical studies, and methods chosen will also vary with the project. Input device research might include some empirical or usability studies, but the novelty of the tools and related tasks often require design explorations with little or no evaluation to simply explore possibilities (Ishii and Ullmer, 1997). In another example, a project involving the development of digital family portraits used preliminary interviews to assess needs for a proposed tool. This formative work was followed by field studies combining use of the tool in participants' homes with ongoing diary and interview studies. These later studies provided detailed insight into how one family made use of the tools (Mynatt et al., 2001; Rowan and Mynatt, 2005).

The importance of contextual factors and the novelty of the technologies involved make up-front qualitative investigations vitally important, as misunderstandings of contexts and how tools might be used can often compromise the goals of the system. Interviews with users, particularly when conducted in context (Vaisutis et al., 2014; Mynatt et al., 2001; Rowan and Mynatt, 2005) can identify key requirements and guide design. Similarly, field studies are often vital for understanding impacts of the tools in context (Houben et al., 2016), particularly in terms of the many privacy and control concerns associated with unobtrusive sensors and monitors that might capture activities that some individuals might not want recorded (Mynatt et al., 2001; Rowan and Mynatt, 2005; Kärkkäinen et al., 2010). Despite the strengths of these field studies, they remain expensive and challenging, leading some to look for alternative approaches such as scale models useful for inexpensively demonstrating and exploring possibilities (Chatzigiannakis et al., 2014).

Other efforts have explored the use of more quantitative versions of diaries. One exploration looked at the frequencies and locations of interactions with all things in the home—whether digital or not—during intervals lasting several hours. Histograms indicating the frequency of use of various objects, along with maps describing locations, provide detailed understanding of interactions with objects that might conceivably be integrated into an Internet of Things (Crabtree et al., 2006). Similar approaches to studying the use of everyday, noncomputer onjects and how those uses might inform the Internet of Things have led to the development of "object-oriented ethnography"—the study of how objects interact with our lives and how this understanding might inform the design of devices augmented with computing capabilities (Nansen et al., 2014).

Design sessions, similar to those used in participatory design, can also be useful for developing ideas. One effort examined devices, interactions, and roles to develop a set of graphical building blocks describing roles of individuals and devices in various ubiquitous computing scenarios, using these blocks to both convey elements of design and to identify difficulties with proposed designs (Kim et al., 2016b). Although the development of methods of this sort quickly becomes its own research project involving multiple iterations and revisions, the results can often be very informative.

Sensor data presents its own challenges. Inexpensive sensors for detecting motion and sound might be available commercially, either as "ready-to-deploy" products or as components suitable for assembly or control via Arduino or Raspberry PI hardware (Diez and Posada, 2013). Transferring data from sensors to remote servers, through wireless or USB connections will be a requirement. You will also need a plan for storing or processing sensor data, which can be quite voluminous. A willingness to tinker and to consider outrageous ideas can be helpful in tackling these projects, as unexpected approaches can often be quite helpful. In response to the potential expense and complexity of sensors used in projects like digital family portraits, one study found that microphones taped to pipes in a basement could inexpensively and accurately identify the use of bathrooms (sinks, showers, and toilets) and kitchen appliances (Fogarty et al., 2006).

Like physiological data discussed in Chapter 13, sensor-based ubiquitous computing also requires thoughtful planning and careful analysis including preprocessing, filtering, detection of specific types of signals, classification of activities, and storage and management of data—possibly both raw and processed and related patterns. Researchers have proposed a variety of processing pipelines (Ghose et al., 2013), data monitoring tools (Bannach et al., 2010), reference datasets (Reiss and Stricker, 2012), data integration strategies (Schuldhaus et al., 2014), and database architectures, including the use of so-called NoSQL databases (Zeni et al., 2014), to address these challenges. Details of these components will generally be dependent upon sensor capabilities, requirements for data storage, and the specific questions being asked.

Triangulated coordination of data collection and interpretation can often be highly informative. A study of movement throughout places in the home illustrates the potential for coordination of automated data collection with qualitative data—in this case, interviews (Aipperspach et al., 2006). To understand the patterns of activity in homes, researchers placed location sensors at various points throughout several homes. These sensors captured where people were, when they were there, and for how long. Mathematical models were used to combine individual events in the log files into meaningful aggregates that identified "places"—locations of significant activity in the home. The models were evaluated by comparing the automatically identified places with the results of interviews with the participants. Interviews with the participants had the added advantage of providing context to explain some of the results of the models. In one case, models identified a "place" that included both a kitchen table and a living room couch. Interviews with the residents of this particular home indicated that this data was collected during the course of a birthday party, when they were continually moving between the kitchen and the living room, acting as if the two locations were part of one larger space (Aipperspach et al., 2006). Automated methods that focused only on the contents of the activity logs would not have had access to this more nuanced explanation of resident activity.

14.5 SUMMARY

As the growth of the Internet and the availability of low-cost sensors led information and computing into ubiquitous and familiar roles pervading all aspects of everyday life, it seems only natural that these technologies would play key roles in HCI research. Remote usability studies simplify the process of conducting usability studies while providing access to larger scales of data. Human computation systems have opened the door to an entirely new type of data collection, harnessing the power of networks to engage many individuals in completing small tasks providing insight unavailable through computational tools. Sensor-based systems allow for the easy collection of new types of data in volumes not previously imaginable.

Online activity also provides a rich source of data for close examination of complex interactions and communication patterns. Examination of online discussions and interactions can tell us a great deal about the dynamics of group conversation and the spread of key ideas.

Online and ubiquitous HCI research will likely continue to be shaped by—and to shape—emerging tools and technological approaches. Improved conferencing tools with richer integration of simultaneous screen-sharing and webcam feedback will likely enable richer and more informative online research studies, while new social applications will enable novel interaction patterns, promote further study, and suggest further innovations in the next generation of tools and applications. Comparable advances in sensors and ubiquitous tools will facilitate the collection of richer, higher-resolution, and higher-fidelity data.

As the scope of online HCI research increases, ethical concerns associated with frequent and often unobserved data collection will expand as well. Although the comparison of alternative web site designs via A/B testing may be relatively benign, integrated analyses of social media interactions, health data collected by wearable devices, and other ubiquitous sensor data may identify insights not possible from any single dataset, possibly revealing sensitive information that some participants might prefer to leave undiscussed.

DISCUSSION QUESTIONS

1. The earlier discussion of remote usability testing cited a lack of direct feedback from participants as a possible drawback. For in-person studies, careful observation of facial expression and body language during usability tests might help researchers identify moments of frustration or other emotional responses. These cues might not be available in synchronous remote usability studies, as some web conferencing tools might have limited video feedback through webcams, while others support only screen sharing. However, there might be some advantages to the lack of direct feedback. For example, in some cases, participants might be willing to be more frank in providing direct feedback if they do not have to see the facial expressions of the person administering the experiment. Are there other types of studies or questions that might provide better feedback if conducted remotely (as opposed to in-person)? How might you evaluate the suitability of different questions for remote versus in-person usability studies?

2. Ubiquitous mobile social computing through smartphones blurs lines between "traditional" social networks and sensor-based ubiquitous computing. From location-based apps used for identifying friends who might be physically nearby to the Pokemon Go game that challenged users to explore and find Pokemon characters on city streets and in natural environments, these apps present the possibility of capturing and studying very rich datasets. What are some of the challenges associated with studying datasets that might combine geographical locations, social media posts, and detailed interactions with custom apps? Are there additional ethical dilemmas associated with the combination of these various data types?

3. Many crowdsourcing user studies might be seen as generalizations of remote usability studies, conducted through a software platform built to support participant recruitment. However, incentives might differ: in traditional lab studies, participants might be offered some money or a gift for participating, but crowdsourcing workers are generally paid by the task. Does this approach raise any concerns regarding the ethical treatment of research participants?

RESEARCH DESIGN EXERCISE

The combination of human computation and ubiquitous computing raises some interesting and challenging opportunities for HCI research. Imagine a novel application of the intersection of these techniques designed to help with a distinctly ancient and noncomputerized human activity: gardening. Specifically, a gardening support network might use online fora (or is it flora?) for members to exchange information and tips about cultivation of various plants in different climes. Participants might use ubiquitous computing tools to capture photos of plants, to measure activity in watering, and to track time spent working on the garden. Finally, human computation elements might be used to verify the identity of unfamiliar plants or blights or other infections that might harm plants: images collected from an individual's garden might be sent to a community of workers who might theorize about the identity of the plant in question, with a majority vote summarizing the consensus of the community. Speculate as to how you might go about constructing and studying this complex ecosystem. What design issues and challenges do you see? How might issues such as differing levels of expertise and experience be accounted for in the design? How might users distinguish between good advice and bad? How can users be enticed to participate in the interpretation of provided images? How might you evaluate the success of the various components of this system?

REFERENCES

Abdul-Rahman, A., Proctor, K.J., Duffy, B., Chen, M., 2014. Repeated measures design in crowdsourcing-based experiments for visualization. In: Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization. ACM, Paris, pp. 95–102.

Ahn, L.v., Dabbish, L., 2004. Labeling images with a computer game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vienna, pp. 319–326.

Ahn, L.v., Dabbish, L., 2008. Designing games with a purpose. Communications of the ACM 51 (8), 58–67.

Ahn, L.V., Blum, M., Hopper, N.J., Langford, J., 2003. CAPTCHA: using hard AI problems for security. In: Proceedings of the 22nd International Conference on Theory and Applications of Cryptographic Techniques. Springer-Verlag, Warsaw, pp. 294–311.

Aipperspach, R., Rattenbury, T., Woodruff, A., Canny, J., 2006, September. A quantitative method for revealing and comparing places in the home. In: International Conference on Ubiquitous Computing. Springer Berlin Heidelberg, pp. 1–18.

Allen, C., Tsou, M.H., Aslam, A., Nagel, A., Gawron, J.M., 2016. Applying GIS and machine learning methods to twitter data for multiscale surveillance of influenza. PLoS One 11 (7), e0157734.

Amazon, 2016. Amazon Mechanical Turk Documentation. Retrieved August 30, 2016, from https://aws.amazon.com/documentation/mturk/.

Anand, P.K., Matthew, C., Bjoern, H., 2011. Turkomatic: automatic recursive task and workflow design for mechanical Turk. In: CHI '11 Extended Abstracts on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 2053–2058.

Andreasen, M.S., Neilsen, H.V., Schrøder, S.O., Stage, J., 2007. What happened to remote usability testing?: an empirical study of three methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, San Jose, CA, pp. 1405–1414.

Ashton, K., 2009. That 'Internet of Things' thing. RFID Journal. http://www.rfidjournal.com/articles/view?4986 (accessed March 19, 2017).

Azar, B., 2000. Online experiments: ethically fair or foul? Monitor on Psychology 31 (4), 50–51.

Baeza-Yates, R., Riberio-Neto, B., 2011. Modern Information Retrieval: The Concepts and Technology Behind Search. Addison-Wesley-Longman, Reading, MA.

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L., 2012. The role of social networks in information diffusion. In: Proceedings of the 21st International Conference on World Wide Web. ACM, Lyon, pp. 519–528.

Bannach, D., Kunze, K., Weppner, J., Lukowicz, P., 2010. Integrated tool chain for recording and handling large, multimodal context recognition data sets. In: Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing—Adjunct. ACM, Copenhagen, pp. 357–358.

Barowy, D.W., Curtsinger, C., Berger, E., McGregor, A., 2016. AutoMan: a platform for integrating human-based and digital computation. Communications of the ACM 59 (6), 102–109.

Bernstein, M.S., Brandt, J., Miller, R.C., Karger, D.R., 2011. Crowds in two seconds: enabling realtime crowd-powered interfaces. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, Santa Barbara, CA, pp. 33–42.

Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K., 2015. Soylent: a word processor with a crowd inside. Communications of the ACM 58 (8), 85–94.

Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, R.C., Miller, R., Miller, A., Tatarowicz, B., White, S., White, T., Yeh, T., Miller, R., 2010. VizWiz: nearly real-time answers to visual questions. In: Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, pp. 333–342.

Bonsignore, E.M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D.L., Shneiderman, B., 2009. First steps to Netviz Nirvana: evaluating social network analysis with NodeXL. In: International Conference on Computational Science and Engineering, CSE '09, pp. 332–339.

Bruckman, A.s., 2002. Ethical Guidelines for Research Online. Retrieved March 19, 2017, from http://www.cc.gatech.edu/~asb/ethics/.

Brush, A.J., Ames, M., Davis, J., 2004. A comparison of synchronous remote and local usability studies for an expert interface. Extended Abstracts, ACM Conference on Human Factors in Computing Systems, ACM Press, Vienna, pp. 1179–1182.

Bruun, A., Stage, J., 2012. The effect of task assignments and instruction types on remote asynchronous usability testing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Austin, TX, pp. 2117–2126.

Bruun, A., Gull, P., Hofmeister, L., Stage, J., 2009. Let your users do the testing: a comparison of three remote asynchronous usability testing methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Boston, MA, pp. 1619–1628.

Burke, M., Marlow, C., Lento, T., 2010. Social network activity and social well-being. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Atlanta, GA, pp. 1909–1912.

Bush, V., 1945. As we may think. The atlantic monthly 176 (1), 101–108.

Butler, D., 2013. When Google got flu wrong. Nature 494, 155–156.

Cai, C.J., Iqbal, S.T., Teevan, J., 2016. Chain reactions: the impact of order on microtask chains. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 3143–3154.

Cassa, C.A., Chunara, R., Mandl, K., Brownstein, J.S., 2013. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. PLos Currents Disasters. Edition 1. http://dx.doi.org/10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b.

Centivany, A., Glushko, B., 2016. "Popcorn tastes good": participatory policymaking and Reddit's. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 1126–1137.

Chatzigiannakis, I., Drude, J.P., Hasemann, H., Kröller, A., 2014. Developing smart homes using the internet of things: how to demonstrate your system. In: Streitz, N., Markopoulos, P. (Eds.), Distributed, Ambient, and Pervasive Interactions: Second International Conference, DAPI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014. Springer International Publishing, pp. 415–426.

Chen, T.Y.-H., Ravindranath, L., Deng, S., Bahl, P., Balakrishnan, H., 2015. Glimpse: continuous, real-time object recognition on mobile devices. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. ACM, Seoul, pp. 155–168.

Chynał, P., Szymański, J.M., 2011. Remote Usability Testing Using Eyetracking. In: Campos, P., Graham, N., Jorge, J., et al. (Eds.), Human-Computer Interaction—INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5–9, 2011, Part 1. Springer, Berlin, Heidelberg, pp. 356–361.

Cook, S., Conrad, C., Fowlkes, A.L., Mohebbi, M.H., 2011. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 6 (8), e23610.

Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M., Brown, B., 2006. Supporting ethnographic studies of ubiquitous computing in the wild. In: ACM Conference on Designing Interactive Systems. ACM Press, University Park, PA, pp. 60–69.

Crook, T., Frasca, B., Kohavi, R., Longbotham, R., 2009. Seven pitfalls to avoid when running controlled experiments on the web. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Paris, pp. 1105–1114.

CrowdFlower, 2016. API | CrowdFlower. Retrieved August 30, 2016, 2016, from https://www.crowdflower.com/docs-api/

Dai, P., Rzeszotarski, J.M., Paritosh, P., Chi, E.H., 2015. And now for something completely different: improving crowdsourcing workflows with micro-diversions. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing. ACM, Vancouver, BC, pp. 628–638.

Davidson, M., Renaud, K., Li, S., 2014. jCAPTCHA: accessible human validation. In: Miesenberger, K., Fels, D., Archambault, D., Peňáz, P., Zagler, W. (Eds.), Computers Helping People with Special Needs: 14th International Conference, ICCHP 2014, Paris, France, July 9–11, 2014, Part I. Springer International Publishing, pp. 129–136.

Deng, A., 2015. Objective Bayesian two sample hypothesis testing for online controlled experiments. In: Proceedings of the 24th International Conference on World Wide Web. ACM, Florence, p. 913.

Deng, A., Xu, Y., Kohavi, R., Walker, T., 2013. Improving the sensitivity of online controlled experiments by utilizing pre-experiment data. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. ACM, Rome, pp. 123–132.

Deng, A., Li, T., Guo, Y., 2014. Statistical inference in two-stage online controlled experiments with treatment selection and validation. In: Proceedings of the 23rd International Conference on World Wide Web. ACM, Seoul, Korea, pp. 609–618.

Diez, T., Posada, A., 2013. The fab and the smart city: the use of machines and technology for the city production by its citizens. In: Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction. ACM, Barcelona, pp. 447–454.

Dray, S., Siegel, D., 2004. Remote possibilities?: international usability testing at a distance. Interactions 11 (2), 10–17.

Eiben, C.B., Siegel, J.B., Bale, J.B., Cooper, S., Khatib, F., Shen, B.W., Players, F., Stoddard, B.L., Popovic, Z., Baker, D., 2012. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. Nature Biotechnology 30 (2), 190–192.

Embiricos, A., Rahmati, N., Zhu, N., Bernstein, M.S., 2014. Structured handoffs in expert crowdsourcing improve communication and work output. In: Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology. ACM, Honolulu, Hawaii, pp. 99–100.

Engelbart, D.C., 1962. Augmenting Human Intellect: A Conceptual Framework. SRI Summary Report.

Filippova, A., Cho, H., 2016. The effects and antecedents of conflict in free and open source software development. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM, San Francisco, CA, pp. 705–716.

Finley, K., 2014. Twitter Opens Its Enormous Archives to Data-Hungry Academics. Wired. Retrieved August 20, 2016, from http://www.wired.com/2014/02/twitter-promises-share-secrets-academia/.

Fogarty, J., Au, C., Hudson, S.E., 2006. Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology. ACM, Montreux, pp. 91–100.

Frankel, M.S., Siang, S., 1999. Ethical and Legal Aspects of Human Subjects Research on the Internet. American Association for the Advancement of Science.

Ghose, A., Sinha, P., Bhaumik, C., Sinha, A., Agrawal, A., Choudhury, A.D., 2013. UbiHeld: ubiquitous healthcare monitoring system for elderly and chronic patients. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. ACM, Zurich, pp. 1255–1264.

Gilbert, E., 2012. Phrases that signal workplace hierarchy. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. ACM, Seattle, WA, pp. 1037–1046.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457 (7232), 1012–1014.

Good, B.M., Loguercio, S., Nanis, M., Su, A.I., 2012. Genegames.org: high-throughput access to biological knowledge and reasoning through online games. In: IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB), p. 145.

Greg, L., Lydia, B.C., Max, G., Robert, C.M., 2010. TurKit: human computation algorithms on mechanical Turk. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, ISBN: 978-1-4503-0271-5, pp. 57–66.

Hansen, D., Shneiderman, B., 2010. Analyzing social media networks with NodeXL: insights from a connected world. Morgan Kaufmann, Burlington, MA.

Harkema, H., Dowling, J.N., Thornblade, T., Chapman, W.W., 2009. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. Journal of Biomedical Informatics 42 (5), 839–851.

Hartson, H.R., Castillo, J.C., Kelso, J., Neale, W.C., 1996. Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 228–235.

Hedegaard, S., Simonsen, J.G., 2013. Extracting usability and user experience information from online user reviews. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Paris, pp. 2089–2098.

Hedegaard, S., Simonsen, J.G., 2014. Mining until it hurts: automatic extraction of usability issues from online reviews compared to traditional usability evaluation. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. ACM, Helsinki, pp. 157–166.

Heer, J., Bostock, M., 2010. Crowdsourcing graphical perception: using mechanical Turk to assess visualization design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Atlanta, GA, pp. 203–212.

Hochheiser, H., Ning, Y., Hernandez, A., Horn, J.R., Jacobson, R., Boyce, R.D., 2016. Using nonexperts for annotating pharmacokinetic drug-drug interaction mentions in product labeling: a feasibility study. JMIR Research Protocols 5 (2), e40.

Houben, S., Golsteijn, C., Gallacher, S., Johnson, R., Bakker, S., Marquardt, N., Capra, L., Rogers, Y., 2016. Physikit: data engagement through physical ambient visualizations in the home. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 1608–1619.

Hu, C., Resnik, P., Bederson, B.B., 2014. Crowdsourced monolingual translation. ACM Transactions on Computer-Human Interaction 21 (4), 1–35.

Hu, G., Chen, L., Okerlund, J., Shaer, O., 2015. Exploring the use of Google Glass in Wet Laboratories. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. ACM, Seoul, pp. 2103–2108.

Huang, J., White, R.W., Dumais, S., 2011. No clicks, no problem: using cursor movements to understand and improve search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 1225–1234.

Huang, J., White, R.W., Buscher, G., Wang, K., 2012. Improving searcher models using mouse cursor activity. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Portland, OR, pp. 195–204.

Huh, J., Liu, L.S., Neogi, T., Inkpen, K., Pratt, W., 2014. Health vlogs as social support for chronic illness management. ACM Transactions on Computer-Human Interaction 21 (4), 23:1–23:31.

Ipeirotis, P.G., Provost, F., Wang, J., 2010. Quality management on amazon mechanical Turk. In: Proceedings of the ACM SIGKDD Workshop on Human Computation. ACM, Washington, DC, pp. 64–67.

Irani, L.C., Silberman, M.S., 2013. Turkopticon: interrupting worker invisibility in amazon mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Paris, ISBN: 978-1-4503-1899-0, pp. 611–620.

Ishii, H., Ullmer, B., 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. ACM, Atlanta, GA, pp. 234–241.

Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, I., Raisamo, R., 2014. Using gaze gestures with haptic feedback on glasses. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. ACM, Helsinki, pp. 1047–1050.

Kärkkäinen, T., Vaittinen, T., Väänänen-Vainio-Mattila, K., 2010. I don't mind being logged, but want to remain in control: a field study of mobile activity and context logging. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Atlanta, GA, pp. 163–172.

Kelly, D., Gyllstrom, K., 2011. An examination of two delivery modes for interactive search system experiments: remote and laboratory. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 1531–1540.

Kharitonov, E., Vorobev, A., Macdonald, C., Serdyukov, P., Ounis, I., 2015. Sequential testing for early stopping of online experiments. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Santiago, Chile, pp. 473–482.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., Jaskolski, M., Baker, D., 2011. Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nature Structural & Molecular Biology 18 (10), 1175–1177.

Kim, J.G., Kong, H.K., Karahalios, K., Fu, W.-T., Hong, H., 2016a. The power of collective endorsements: credibility factors in medical crowdfunding campaigns. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 4538–4549.

Kim, Y., Kim, H., Bae, S.-H., Lee, S., Kim, C.-J., 2016b. Building blocks for designing future multi-device interaction. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 2548–2554.

Kirkorian, R., 2014. Introducing Twitter Data Grants. Retrieved August 20, 2016, from https://blog.twitter.com/2014/introducing-twitter-data-grants.

Kittur, A., Kraut, R.E., 2008. Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work. ACM, San Diego, CA, pp. 37–46.

Kittur, A., Chi, E.H., Suh, B., 2008. Crowdsourcing user studies with mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Florence, pp. 453–456.

Klembczyk, J.J., Jalalpour, M., Levin, S., Washington, R.E., Pines, J.M., Rothman, R.E., Dugas, A.F., 2016. Google Flu trends spatial variability validated against emergency department influenza-related visits. Journal of Medical Internet Research 18 (6), e175.

Kohavi, R., Longbotham, R., Sommerfield, D., Henne, R.M., 2009. Controlled experiments on the web: survey and practical guide. Data Mining and Knowledge Discovery 18 (1), 140–181.

Kohavi, R., Deng, A., Frasca, B., Longbotham, R., Walker, T., Xu, Y., 2012. Trustworthy online controlled experiments: five puzzling outcomes explained. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Beijing, pp. 786–794.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., Pohlmann, N., 2013. Online controlled experiments at large scale. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Chicago, IL, pp. 1168–1176.

Kong, N., Hearst, M.A., Agrawala, M., 2014. Extracting references between text and charts via crowdsourcing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Toronto, ON, pp. 31–40.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., Couper, M., 2004. Psychological research online: report of board of scientific affairs' advisory group on the conduct of research on the internet. The American Psychologist 59 (2), 105–117.

Krishna, R.A., Hata, K., Chen, S., Kravitz, J., Shamma, D.A., Fei-Fei, L., Bernstein, M.S., 2016. Embracing error to enable rapid crowdsourcing. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 3167–3179.

Kucherbaev, P., Daniel, F., Tranquillini, S., Marchese, M., 2016. ReLauncher: crowdsourcing micro-tasks runtime controller. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM, San Francisco, CA, pp. 1609–1614.

Lasecki, W.S., Zhong, Y., Bigham, J.P., 2014. Increasing the bandwidth of crowdsourced visual question answering to better support blind users. In: Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility. ACM, Rochester, NY, pp. 263–264.

Lasecki, W.S., Kim, J., Rafter, N., Sen, O., Bigham, J.P., Bernstein, M.S., 2015. Apparition: crowdsourced user interfaces that come to life as you sketch them. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, Seoul, pp. 1925–1934.

Law, E., Ahn, L.V., 2011. Human computation. Synthesis Lectures on Artificial Intelligence and Machine Learning 5 (3), 1–121.

Law, E., Yin, M., Goh, J., Chen, K., Terry, M.A., Gajos, K.Z., 2016. Curiosity killed the Cat, but makes crowdwork better. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 4098–4110.

Lazar, J., Preece, J., 1999. Designing and implementing web-based surveys. Journal of Computer Information Systems 39 (4), 63–67.

Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The parable of Google flu: traps in big data analysis. Science 343 (6176), 1203–1205.

Leug, C., Fisher, D., 2003. From Usenet to CoWebs: Interacting With Social Information Spaces. Springer, London.

Madathil, K.C., Greenstein, J.S., 2011. Synchronous remote usability testing: a new approach facilitated by virtual worlds. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 2225–2234.

Maloney-Krichmar, D., Preece, J., 2005. A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM Transactions on Computer-Human Interaction 12 (2), 201–232.

Mao, A., Chen, Y., Gajos, K.Z., Parkes, D.C., Procaccia, A.D., Zhang, H., 2012. TurkServer: enabling synchronous and longitudinal online experiments. In: Fourth Workshop on Human Computation (HCOMP 2012).

Matias, J.N., 2016. Going dark: social factors in collective action against platform operators in the reddit blackout. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 1138–1151.

McInnis, B., Leshed, G., 2016. Running user studies with crowd workers. Interactions 23 (5), 50–53.

McNaney, R., Vines, J., Roggen, D., Balaam, M., Zhang, P., Poliakov, I., Olivier, P., 2014. Exploring the acceptability of Google glass as an everyday assistive device for people with Parkinson's. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Toronto, ON, pp. 2551–2554.

Menefee, H.K., Thompson, M.J., Guterbock, T.M., Williams, I.C., Valdez, R.S., 2016. Mechanisms of communicating health information through Facebook: implications for consumer health information technology design. Journal of Medical Internet Research 18 (8), e218.

Meyer, J., Bederson, B., 1998. Does a Sketchy Appearance Influence Drawing Behavior? Human Computer Interaction Lab Tech Report. University of Maryland. Retrieved March 20, 2017 from http://hcil2.cs.umd.edu/trs/98-12/98-12.pdf.

Micallef, L., Dragicevic, P., Fekete, J.D., 2012. Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing. IEEE Transactions on Visualization and Computer Graphics 18 (12), 2536–2545.

Michelucci, P., Shanley, L., Dickinson, J., Hirsh, H., 2015. A U.S. Research Roadmap for Human Computation. Computing Community Consortium Technical Report. Available at https://arxiv.org/ftp/arxiv/papers/1505/1505.07096.pdf (accessed April 8, 2017).

Morris, R.R., Schueller, S.M., Picard, R.W., 2015. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. Journal of Medical Internet Research 17 (3), e72.

Mynatt, E.D., Rowan, J., Craighill, S., Jacobs, A., 2001. Digital family portraits: supporting peace of mind for extended family members. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Seattle, WA, pp. 333–340.

Myslín, M., Zhu, S.-H., Chapman, W., Conway, M., 2013. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. Journal of Medical Internet Research 15 (8), e174.

Nansen, B., Ryn, L.v., Vetere, F., Robertson, T., Brereton, M., Dourish, P., 2014. An internet of social things. In: Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design. ACM, Sydney, NSW, pp. 87–96.

Nebeling, M., Speicher, M., Grossniklaus, M., Norrie, M.C., 2012. Crowdsourced web site evaluation with crowdstudy. In: Proceedings of the 12th International Conference on Web Engineering. Springer-Verlag, Berlin, pp. 494–497.

Nebeling, M., Speicher, M., Norrie, M.C., 2013. CrowdStudy: general toolkit for crowdsourced evaluation of web interfaces. In: Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. ACM, London, pp. 255–264.

Nguyen, E., Modak, T., Dias, E., Yu, Y., Huang, L., 2014. Fitnamo: using bodydata to encourage exercise through Google glass. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. ACM, Toronto, Ontario, Canada, pp. 239–244.

Nichols, J., Mahmud, J., Drews, C., 2012. Summarizing sporting events using twitter. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces. ACM, Lisbon, pp. 189–198.

Nielsen, J., 2005. Putting A/B Testing in Its Place. Retrieved July 18, 2016, from https://www.nngroup.com/articles/putting-ab-testing-in-its-place/.

Nielsen, J., 2012. A/B Testing, Usability Engineering, Radical Innovation: What Pays Best? Retrieved July 18, 2016, from https://www.nngroup.com/articles/ab-testing-usability-engineering/.

Nielsen, J., 2014. Define Stronger A/B Test Variations Through UX Research. Retrieved July 18, 2016, from https://www.nngroup.com/articles/ab-testing-and-ux-research/.

Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on Amazon Mechanical Turk. Judgment and Decision making 5 (5), 411–419.

Paparrizos, J., White, R.W., Horvitz, E., 2016. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. Journal of Oncology Practice 12 (8), 737–744.

Petrie, H., Hamilton, F., King, N., Pavan, P., 2006, April. Remote usability evaluations with disabled people. In Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM, pp. 1133–1141.

Quinn, A.J., Bederson, B.B., 2011. Human computation: a survey and taxonomy of a growing field. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, pp. 1403–1412.

Reiss, A., Stricker, D., 2012. Creating and benchmarking a new dataset for physical activity monitoring. In: Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments. ACM, Heraklion, Crete, pp. 1–8.

Richard, G.T., Kafai, Y.B., 2016. Blind spots in youth DIY programming: examining diversity in creators, content, and comments within the scratch online community. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 1473–1485.

Rowan, J., Mynatt, E.D., 2005. Digital family portrait field trial: support for aging in place. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Portland, OR, pp. 521–530.

Rzeszotarski, J.M., Kittur, A., 2011. Instrumenting the crowd: using implicit behavioral measures to predict task performance. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, Santa Barbara, CA, pp. 13–22.

Santillana, M., Nguyen, A.T., Dredze, M., Paul, M.J., Nsoesie, E.O., Brownstein, J.S., 2015. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Computational Biology 11 (10), e1004513.

Sauer, G., Lazar, J., Hochheiser, H., Feng, J., 2010. Towards a universally usable human interaction proof: evaluation of task completion strategies. ACM Transactions on Accessible Computing 2 (4), 1–32.

Schneider, O.S., Seifi, H., Kashani, S., Chun, M., MacLean, K.E., 2016. HapTurk: crowdsourcing affective ratings of vibrotactile icons. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, Santa Clara, CA, pp. 3248–3260.

Schnepp, J., Shiver, B., 2011. Improving deaf accessibility in remote usability testing. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, Dundee, Scotland, pp. 255–256.

Schuldhaus, D., Leutheuser, H., Eskofier, B.M., 2014. Towards big data for activity recognition: a novel database fusion strategy. In: Proceedings of the 9th International Conference on Body Area Networks. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), London, pp. 97–103.

Scott, J., 2013. Social Network Analysis. Sage Publications, Ltd., London.

Shet, V., 2014. Are you a robot? Introducing "No CAPTCHA reCAPTCHA". Google Security Blog. Retrieved August 24, 2016, from https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html.

Shneiderman, B., 2002. Leonardo's Laptop: Human Needs and the New Computing Technologies. MIT Press, Cambridge, MA.

Simoens, P., Coninck, E.D., Vervust, T., Wijmeersch, J.-F.V., Ingelbinck, T., Verbelen, T., Beeck, M.O.d., Dhoedt, B., 2014. Vision: smart home control with head-mounted sensors for vision and brain activity. In: Proceedings of the Fifth International Workshop on Mobile Cloud Computing and Services. ACM, Bretton Woods, NH, pp. 29–33.

Sörös, G., Daiber, F., Weller, T., 2013. Cyclo: a personal bike coach through the glass. In: SIGGRAPH Asia, Symposium on Mobile Graphics and Interactive Applications. ACM, Hong Kong, pp. 1–4.

Starbird, K., Palen, L., 2012. (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. ACM, Seattle, WA, pp. 7–16.
Vaish, R., Wyngarden, K., Chen, J., Cheung, B., Bernstein, M.S., 2014. Twitch crowdsourcing: crowd contributions in short bursts of time. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Toronto, ON, pp. 3645–3654.
Vaisutis, K., Brereton, M., Robertson, T., Vetere, F., Durick, J., Nansen, B., Buys, L., 2014. Invisible connections: investigating older people's emotions and social relations around objects. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems. ACM, Toronto, ON, pp. 1937–1940.
Varnhagen, C.K., Gushta, M., Daniels, J., Peters, T.C., Parmar, N., Law, D., Hirsch, R., Takach, B.S., Johnson, T., 2005. How informed is online informed consent? Ethics & Behavior 15 (1), 37–48.
Viégas, F.B., Wattenberg, M., Dave, K., 2004. Studying cooperation and conflict between authors with *history flow* visualizations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vienna, pp. 575–582.
Viégas, F.B., Wattenberg, M., Kriss, J., Ham, F.V., 2007a. Talk before you type: coordination in Wikipedia. In: 40th Annual Hawaii International Conference on System Sciences, 2007, HICSS 2007, pp. 1–10.
Viégas, F.B., Wattenberg, M., McKeon, M.M., 2007b. The hidden order of Wikipedia. In: Schuler, D. (Ed.), Online Communities and Social Computing: Second International Conference, OCSC 2007, Held as Part of HCI International 2007, Beijing, China, July 22–27, 2007. Springer, Berlin, Heidelberg, pp. 445–454.
von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M., 2008. ReCAPTCHA: human-based character recognition via web security measures. Science 321, 1465–1468.
Weiser, M., 1991. The computer for the 21st century. Scientific american 265 (3), 94–104.
Weld, D.S., 2015. Intelligent control of crowdsourcing. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. ACM, Atlanta, GA, pp. 1–2.
Weppner, J., Hirth, M., Kuhn, J., Lukowicz, P., 2014a. Physics education with Google glass *gPhysics* experiment app. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, Seattle, WA, pp. 279–282.
Weppner, J., Poxrucker, A., Lukowicz, P., Ishimaru, S., Kunze, K., Kise, K., 2014b. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, Seattle, WA, pp. 283–286.
White, R., 2013. Beliefs and biases in web search. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Dublin, pp. 3–12.
White, R.W., Hassan, A., 2014. Content bias in online health search. ACM Transactions on the Web 8 (4), 1–33.
White, R.W., Horvitz, E., 2009. Cyberchondria: studies of the escalation of medical concerns in Web search. ACM Transactions on Information Systems 27 (4), 1–37.
White, R.W., Tatonetti, N.P., Shah, N.H., Altman, R.B., Horvitz, E., 2013. Web-scale pharmacovigilance: listening to signals from the crowd. Journal of the American Medical Informatics Association 20 (3), 404–408.
Winandy, M., Kostkova, P., de Quincey, E., St. Louis, C., Szomszor, M., 2016. Follow #eHealth 2011: measuring the role and effectiveness of online and social media in increasing the outreach of a scientific conference. Journal of Medical Internet Research 18 (7), e191.

Woo, H., Cho, Y., Shim, E., Lee, J.-K., Lee, C.-G., Kim, S.H., 2016. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. Journal of Medical Internet Research 18 (7), e177.

Xiang, R., Neville, J., Rogati, M., 2010. Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World wide Web. ACM, Raleigh, NC, pp. 981–990.

Yang, C., Harkreader, R., Zhang, J., Shin, S., Gu, G., 2012. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: Proceedings of the 21st International Conference on World Wide Web. ACM, Lyon, pp. 71–80.

Zeni, M., Zaihrayeu, I., Giunchiglia, F., 2014. Multi-device activity logging. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, Seattle, WA, pp. 299–302.

Zhang, X., Chen, Z., Fang, C., Liu, Z., 2016. Guiding the crowds for android testing. In: Proceedings of the 38th International Conference on Software Engineering Companion. ACM, Austin, TX, pp. 752–753.

Zhao, R., Wang, J., 2011. Visualizing the research on pervasive and ubiquitous computing. Scientometrics 86 (3), 593–612.

Zhu, S., Kane, S.K., Feng, J., Sears, A., 2012. A crowdsourcing quality control model for tasks distributed in parallel. In: CHI '12 extended abstracts on human factors in computing systems. ACM, Austin, TX, ISBN: 978-1-4503-1016-1, pp. 2501–2506.

Zlatos, B., 2009. Google Purchases Carnegie Mellon University Spin-Off ReCaptcha. Pittsburgh Tribune-Review.