Crowdsourcing as a Tool for Research: Implications of Uncertainty

Edith Law
University of Waterloo
Waterloo ON, Canada
edith.law@uwaterloo.ca

Krzysztof Z. Gajos
Harvard University
Cambridge MA, USA
kgajos@eecs.harvard.edu

Andrea Wiggins
University of Maryland
College Park MA, USA
wiggins@umd.edu

Mary L Gray
Microsoft Research
Cambridge MA, USA
mlg@microsoft.com

Alex Williams
University of Waterloo
Waterloo ON, Canada
alex.williams@uwaterloo.ca

ABSTRACT
Numerous crowdsourcing platforms are now available to support research as well as commercial goals. However, crowdsourcing is not yet widely adopted by researchers for generating, processing or analyzing research data. This study develops a deeper understanding of the circumstances under which crowdsourcing is a useful, feasible or desirable tool for research, as well as the factors that may influence researchers' decisions around adopting crowdsourcing technology. We conducted semi-structured interviews with 18 researchers in diverse disciplines, spanning the humanities and sciences, to illuminate how research norms and practitioners' dispositions were related to uncertainties around research processes, data, knowledge, delegation and quality. The paper concludes with a discussion of the design implications for future crowdsourcing systems to support research.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
crowdsourcing for research; citizen science; interviews

INTRODUCTION
Crowdsourcing is a nascent tool for streamlining the process of gathering, processing and analyzing research data in many fields. Tasks that were previously conducted by a small team of researchers can now be parallelized and processed by millions of volunteers over the Web, making questions that seemed previously impossible now tractable. Crowdsourcing marketplaces (*e.g.*, Amazon Mechanical Turk) are becoming a prevalent tool for researchers in certain disciplines (*e.g.*, political science [1], psychology [5, 54]). Several large-scale *citizen science* platforms (*e.g.*, Zooniverse [51, 78], eBird [92]) attract hundreds of thousands of volunteers offering their time for science and humanities research; literally hundreds of smaller citizen science projects are flourishing [73, 87]. In digital humanities [60], numerous projects elicit the help of volunteers to transcribe manuscripts, enhance metadata, add contextual knowledge to artifacts, co-curate exhibits, etc.

CSCW '17, February 25-March 01, 2017, Portland, OR, USA
©2017 ACM. ISBN 978-1-4503-4335-0/17/03...$15.00
DOI: http://dx.doi.org/10.1145/2998181.2998197

Despite its increasing popularity and capacity for gathering data and enabling collaboration between researchers and the general public, crowdsourcing has not made its way into mainstream research methodologies. Yet, our understanding of *how and when one would use crowdsourcing as a methodological tool for a particular research project* is limited.

To better understand researchers' current perceptions, we investigated the following questions:

- Under what circumstances is crowdsourcing a feasible, desirable, or useful tool for researchers?
- Under what circumstances is crowdsourcing *not* suitable for research?

We analyzed researchers' practices, norms and values to understand how scientific culture and practices mediate the fit of crowdsourcing strategies to scholarly research needs. This paper is not a philosophical or ethical discussion about the merits of crowdsourcing in academic research; instead, it serves as a formative study about researchers as potential users of crowdsourcing systems. We argue that examining the non-technical aspects of knowledge production is a necessary first step in designing crowdsourcing systems that address the particular needs of researchers.

There are many definitions of the term *crowdsourcing*. Existing research-oriented crowdsourcing projects vary in organizational structure and scale, from small, in-person groups collecting observations in the field, to massive, anonymous crowds annotating, collecting or analyzing data using web-based technology. For the purpose of this work, our analysis primarily focuses on revealing design challenges for *technology-mediated participation*, as our goal is to understand how crowdsourcing can help to massively scale up and streamline the research pipeline. We define crowdsourcing as "efforts that engage large numbers of people over the Web to help collect and process data," which are distinct from the small-team, apprenticeship-like interactions that are currently common in research settings.

We conducted in-depth semi-structured interviews with 18 professors at an elite R1 university, representing a variety of academic disciplines across the sciences and humanities. The interviews focused on understanding how data is handled in the entire research cycle from question formulation to analysis; current research practices around delegation, quality control, and data sharing; and researchers' knowledge and perceptions of crowdsourcing technologies. We asked each researcher to tell us a specific *research story*, while avoiding any discussions of crowdsourcing until the end of the conversation. Our core assumption was that the attitudes and values reflected in current research practices could reveal what might make researchers more or less comfortable with crowdsourcing. We believe that research-oriented crowdsourcing platforms need to be designed with these concerns explicitly in mind.

A strong theme that emerged from our interview data was that constant *uncertainties* researchers faced in their everyday processes affected both the activities that they chose to engage in and the tools that they selected to facilitate inquiry. The term "uncertainty" as used here refers to a core concept from management literature on organizations that conceptualizes information processing as "reduction of uncertainty" [33, 83]. This concept was also important in early research on decision-making under conditions of risk and uncertainty (*e.g.*, [84]). Throughout the paper, we use the term *uncertainty* to mean gaps in information needed for decision-making during the process of designing and executing a research project. In the context of scholarly research, uncertainty influenced how researchers navigated the multi-faceted challenges posed by their work, chose strategies to address these challenges, and understood current crowdsourcing technologies' capacities to accommodate these considerations.

In this paper, we first review the literature to frame our analysis and present our research methods. We then organize the findings around five different types of uncertainties that researchers face—process, data, knowledge, delegation and quality—and discuss how these uncertainties relate to the feasibility, desirability and usefulness of crowdsourcing as a tool for research. Table 1, which we will revisit throughout the paper, summarizes the key dimensions for evaluating the suitability of crowdsourcing as a research tool that emerged from our interview data. Finally, we summarize the contributions to CSCW, the key implications for optimizing the design of crowdsourcing technologies to meet researchers' needs, and identify implications for practice, policy, and future research.

	Suits crowds ✓	Less suitable ✗
Process Uncertainties		
research goals	established	changing
workflow	easily decomposable	not decomposable
processing and analysis	separable	inseparable
source of serendipity	diverse perspectives	deep knowledge
Data Uncertainties		
sensitivity	shareable	private
typicality	common	rare
quantity	abundant	scarce
availability	easily accessible	difficult to obtain
research questions supported	many	few
sharing norms	established	not standard practice
Knowledge Uncertainties		
requirements	explicit	implicit
availability	adequately distributed	limited to professionals
Delegation Uncertainties		
crowd interest	interested	disinterested
crowd intention	well-intentioned	malicious
crowd ability	capable	not capable
role of researcher	manager	educator
Quality Uncertainties		
tolerance for ambiguity	high	low
goals	quantity, speed, diversity, coverage, precision	precision only

Table 1. Types of research attitudes and challenges that make crowdsourcing feasible, desirable, useful (✓) versus not (✗)

RELATED WORK

While numerous areas of literature can illuminate this topic of inquiry, we focused on prior work that discussed crowdsourcing in research contexts and the adoption of crowdsourcing technologies, with a brief discussion of studies of research practices and knowledge production.

Crowdsourcing for Research

In the scientific domain, the term *citizen science* has been broadly adopted to refer to crowdsourcing projects where volunteers help generate and process research data. Such projects have a long history before the internet, dating back to the 1900s [37]. This term has since been broadened and used synonymously to refer to crowdsourcing efforts aimed at facilitating academic research in a variety of disciplines, including humanities and social sciences. Today, there are hundreds (if not thousands) of citizen science projects; some are small in scale and initiated by local volunteers (*e.g.*, ReClam the Bay [88]), while others have massive contributor bases (*e.g.*, Zooniverse [78]) and are structured top-down by researchers. Across different types of citizen science projects, researchers' roles vary [76, 89, 64]: they act as contractors (interested parties ask researchers to conduct a specific scientific inquiry), leaders (researchers design projects for which volunteers contribute data), collaborators (researchers design projects for which volunteers contribute data but also help with other parts of the research process), co-creators (researchers design projects, but volunteers participate in most or all aspect of the research process) and colleagues (volunteers conduct research independently).

In recent years, a number of citizen science projects (*e.g.*, GalaxyZoo [51], eBird [92]) have demonstrated the practicality of crowdsourcing for large-scale data collection and annotation tasks. The internet makes it increasingly feasible to connect researchers to millions of participants over the Web, leading to distributed science inquiry at unprecedented scale. For example, the Christmas Bird Count, dating back to 1900, first involved 27 participants from 25 locations in North America. Today, eBird [92] attracts hundreds of thousands of birders contributing millions of observations on a monthly basis, with contributions from every country in the world. In June 2015, Simpson et. al [78] reported that the Zooniverse

[78], a citizen science platform which currently houses around forty projects, has enrolled more than 1.1 million participants, with the number continuously rising.

Similarly, volunteer labor is not new to the arts and humanities; Oomen et. al [60] noted that cultural heritage institutions involved volunteers in curatorial tasks well before the age of the internet. Proctor et. al [63] underscored this notion by suggesting that, despite being associated with the internet, crowdsourcing was used in such organizations as far back as the nineteenth century. However, with the emergence of crowdsourcing as a cyberinfrastructure, the accessibility of both volunteer and paid labor is greater than ever before. Digital humanities scholarship offers numerous online crowdsourcing projects dedicated to the transcription of ancient manuscripts, *e.g.*, Ancient Lives [91], Old Weather [96], AnnoTate [95], Transcribe Bentham [85], Operation War Diary [97], etc. Similar to citizen science, technology enables crowd-powered manuscript processing to generate new sources of research data at an unprecedented speed and scale.

Researchers who want to use crowdsourcing to gather and process data for research can select from a variety of existing online resources, such as EpiCollect.net [41], Citsci.org [18] and Project Noah [56], that provide guidelines and tools for those interested in implementing their own crowdsourced projects. Recently, several DIY platforms have emerged, *e.g.*, Crowdcrafting and PyBossa framework [71, 72], Sensr [46], CrowdCurio [50], Lab in the Wild [66], and Panoptes [9], which aim to lower the barrier of entry to crowdsourcing by providing plug-and-play functionalities for creating crowdsourcing projects. Finally, there are a wide range of commercial crowdsourcing platforms (*e.g.*, Amazon Mechanical Turk) that researchers can use.

Adoption of Crowdsourcing Technology

The boom in crowdsourcing tools, mentioned above, has generated a growing body of work analyzing existing citizen science projects [15, 76, 89, 90, 25, 39, 77, 28]. These studies examine the characteristics of contributors [69] and their motivation for participating [7, 19, 52, 31, 65], the relationship and communication patterns between contributors and researchers [21], and design guidelines for these platforms [81]. Fewer studies, however, have explored researchers' perceptions of crowdsourcing, its applicability to research practices more generally, and the features that could make crowdsourcing platforms more attractive and useful to a greater number of scholars in the humanities and sciences.

Despite the availability of these new crowdsourcing tools for research, studies find that researchers are often hesitant to adopt crowdsourcing technology. Riesch et al [68], for example, conducted interviews with 30 UK scientists and studied their perceptions of and struggles with volunteer-powered research. The study reported that researchers' negative perceptions were driven by both unrealistic expectations of what crowdsourcing can offer and their need to persuade potential reviewers, the wider scientific community, and professional colleagues of the effectiveness of crowdsourcing. They identified ethical issues, such as the reliance on unpaid individuals, public access to raw data, and the concerns about risks for junior scientists, as sources of scientists' skepticism toward crowd-driven research.

More recently, Schlagwein et al. [24] organized a series of focus groups with 28 business scholars from the Asia-Pacific region to understand the requirements of a research-oriented crowdsourcing system. Despite focusing primarily on the functional system design, the study detailed significant factors, such as interface usability and learning "yet another new technology platform," that contributed to negative perceptions of crowdsourcing. Most of the study's respondents agreed that it was both difficult and unethical to rely on volunteers, suggesting a mandatory integrated payment system. However, economics and psychology research have shown that offering payments can also reduce the quantity and quality of intrinsically motivated contributions [32].

Burgess et al. [10] surveyed 423 professional biodiversity scientists and 125 citizen science project managers to understand barriers to crowdsourcing in biodiversity research. They identified four main barriers, including 1) scientists' limited awareness of citizen science projects that align with their needs, 2) bias towards professionally collected data sources, 3) the fact that citizen science is not universally suitable for all biodiversity research, as well as 4) inconsistency in the quality of data arising from citizen science projects.

Shirk [75] conducted narrative research with 9 scientists who led citizen science projects in the area of conservation, had already adopted crowdsourcing, and were at a point where the success of their research depended on the sustainability of their projects. Shirk focused on understanding the ways they implemented and persisted in conducting their work with volunteers. Her study showed that the scientists who engaged crowds in their research had to carefully negotiate perceptions amongst peers and superiors, constantly demonstrating that they were doing real science and not public service. One of Shirk's subjects characterized her choice to "pursue research more aligned with public interests and social action than with publication" as "difficult but intentional," preferring to reframe her research than to deal with "others' perceptions that their involvement with the public is in conflict with ... knowledge production." In other words, some researchers believed that they could do cutting edge research with crowdsourcing, but at the cost of derision of their peers and little or no recognition.

We similarly elicited perspectives from researchers to "give voice to things that are implicitly known but seldom perceived as important and therefore seldom discussed" [75]. Our work complements Shirk's focus on researchers overcoming challenges and implementing crowdsourcing projects by examining the anticipated needs and concerns of researchers *across disciplines* who have *not* fully adopted crowdsourcing as a research tool. We now turn to the foundational sociological studies of science to frame why a researcher's sense of the meaning and practice of scholarly inquiry might shape the adoption or rejection of crowdsourcing.

Study of Research Practices and Knowledge Production

While the focus of our work was not to provide a detailed sociological account of researchers and how they conduct academic work, there are rich bodies of literature in Sociology of Scientific Knowledge (SSK) and Science and Technology Studies (STS) that examine the cultures of scientific research and the process of scientific discovery [47, 13, 34, 74]. For example, Gieryn's foundational study illustrated how pioneering 19th-century British scientists deployed claims of professional status and objectivity to demarcate their work [35] from amateur contributions. Studies of particle physicists upended depictions of solitary masterminds with preternatural mathematical skills discovering dark matter, showing instead the deep collaborations among the field's lead scientists to advance our knowledge of the universe [82]. Lastly, research analyzing the practices of Salk Institute bench scientists revealed their reliance on annotators who wrote up research results [49]. These are all examples of how social practices in science shaped when and how researchers recognized legitimate contributions to scientific inquiry. As Shapin noted, studies of scientific practice reveal that science is the product of researchers doing a job, not just the uncovering of replicable, generalizable, universal truths [74]. As such, concerns about professional legitimacy, expertise, and whether one is recognizable to disciplinary gatekeepers all impact the generation of scientific knowledge.

The SSK and STS literatures show that scientific inquiry is, first and foremost, an expansive conversation. Scientific collaboration is no longer confined to small groups in laboratories, but also takes place over the internet between large, geographically dispersed teams [38]. Scientific collaboratories [4, 58, 59], for example, are cyberinfrastructures that bridge the gap between scientists "disparate in both geographical location and disciplinary background" [45], providing expert communities with tools to stay constantly connected. Many concerns raised by the informants in our study (*e.g.*, premature exposure of research data and intent) have been well studied in the scientific collaboratories literature [2, 3, 23, 93, 94], though not in the context of their impact on the adoption of crowdsourcing and related technologies. Large cyberinfrastructures for coordinating scientific inquiry can now include non-experts who may only spend a small amount of time on a project collaborating for personal fulfillment and growth [51, 6]. Leaving aside the considerations of professional rank and pay, this is not so different from massive-scale collaborations (*e.g.*, the Large Hadron Collider) that involve numerous individuals filling a wide variety of roles not traditionally associated with scientific inquiry tasks as such.

STUDY DESIGN

In this research, we sought to discover the characteristics of the research processes, challenges, strategies and values of researchers from vastly different fields. We obtained first-hand accounts of current research practices to understand the potential motivations and deterrents to adopting crowdsourcing for academic inquiry. We interviewed 18 researchers, recruited via snowball sampling, whose backgrounds spanned a range of academic disciplines – namely 1 political science, 1 astronomy, 1 palaeontology, 2 neurology, 1 ecology, 1 sociology, 3 biology, 3 history, 2 psychology, 1 physics, 1 chemistry and 1 archaeology – as well as academic ranks, with 3 assistant, 3 associate, 11 full, and 1 retired professor. Intentionally, demographic information is reported only in aggregates in order to preserve respondents' privacy and anonymity. The variability among our research participants reflected a range of organizational values across and within different academic disciplines. Our goal was not to exhaustively capture all possible scenarios or to map disciplinary differences, but to identify emergent patterns among the diverse perspectives and practices.

Each interview lasted 1–2 hours and was conducted at participants' offices, where they could easily access and show examples of their data. At the beginning of each interview, we explained that our goal was to learn how researchers currently conducted their inquiries. We expected that most researchers would have little knowledge of crowdsourcing, and discussions of crowdsourcing would therefore be hypothetical rather than grounded in experience. Indeed, we found that the majority of our interviewees had only a vague understanding of crowdsourcing. Only five interviewees had limited knowledge of crowdsourcing, *i.e.*, awareness of crowdsourcing projects and platforms beyond their own disciplines, and four had used some form of crowdsourcing in their research.

We asked participants to focus on a specific research project, and to describe the research questions and processes through which data were collected, transformed, and analyzed. The interviews were structured around a set of questions about their research process, invoking the themes of expertise, quality control, quality-quantity tradeoffs, efficiency, management of students (*i.e.*, recruiting, training and monitoring), data sharing, and collaboration. Only towards the end, in the last 5–10 minutes of the session, did we probe the interviewees' knowledge of and experience with crowdsourcing. Table 2 summarizes the interview questions in sequence.

Analysis

The audio recordings of the interviews were transcribed and analyzed through iterative inductive content analysis [20]. During the first cycle of coding, we classified excerpts into several iteratively-refined hierarchical codes and developed a coding schema.

The primary and secondary codes included:

- **Data** Properties of the raw data: quantity needed and available, noise, accessibility;
- **Task** Properties of the research task: decomposability, time requirements, teachability, ambiguity, repeatability;
- **Activity** Aspects of the research task: effort, engagement, expertise requirement;
- **Product** Properties of what was produced: types of knowledge contribution, error tolerance;
- **Style** Properties of how research was conducted: distance from data, workflow style, attitude towards new methodologies, use of automation, approaches to quality control, delegation, collaboration with expert peers, universality of research methods, ways to cope with expectations and external pressures;

1	Focus on a particular research question that you studied, can you walk me through the different phases of how data is collected and processed in order to answer that question?
2	What expertise or subjectivity is involved with the task?
3	Who collect, annotate or analyze the data? How are they trained or supervised?
4	If the data is collected by someone else, to what extent can many research questions be answered using the same annotations? Or does the data need to be annotated in a different way each time?
5	Can you think of a scenario for non-expert (*e.g.*, someone without scientific background) to help collect or analyze some type of data that does not require as much expertise, but which still generates useful information that can help the researchers reach their answer?
6	How often do you in the situation where you need to consult experts from a different institution or field? How do you go about finding these people?
7	How do you ensure that the data is collected, annotated or analyzed correctly?
8	What types of errors do you report in publications? Do you report measurement errors?
9	How much do you care about the quality versus the quantity of data?
10	How would it change the way you do science if you had 100x more data?
11	Can you think of situations where you might make use of large amount of lower quality annotations as opposed to small amount of high quality annotations?
12	How would it change the way you do science if you can have the results back 100x faster?
13	Is the raw/annotated data reused (*e.g.*, published? publicly accessible)? If so, how are they shared? For what purpose (*e.g.*, for others to replicate experiments)?
14	Are there particular worries about having data seen by other scientists or people?
15	Do you know what the words "crowdsourcing", "human computation" or "citizen science" mean?
16	Can you describe some existing technology that you know of in each category?
17	How willing do you think an average person / student is motivated or interested in collecting or analyzing the type of data that you work with?
18	Can you think of any legal limitations (*e.g.*, due to privacy issues) or physically limitations (*e.g.*, due to lack of access to proper tools/software/equipment) that will make it difficult for people to help collect or analyze the data?

Table 2. The sequence of questions used to guide the conversation during the interview. The questions (1-14) above the dotted line were designed to guide the researchers in telling their *research story*, while the questions below the dotted line (15-18) discuss perception and knowledge of crowdsourcing.

- **Sharing** Properties of how research was shared: attitude towards sharing intermediate research materials, willingness to share, methods for sharing, cost of sharing;
- **Crowdsourcing** Properties of crowd-based research: knowledge of crowd-powered tools, expectations of crowd attitude, ability and interest, perceptions of crowdsourcing as a tool for research versus service to the general public.

In the first pass analysis, multiple researchers coded statements from interview transcripts generating the primary and secondary codes, then explored the relationships amongst the codes using affinity diagrams. This analysis revealed the pervasive need to manage uncertainties in the research process as a strong theme in all the research stories. The second cycle of coding identified the five types of uncertainties in research based on recurrent themes in the coded data, which seem to directly influence how feasible, desirable or useful researchers believed crowdsourcing to be as a research tool. Our findings are organized around these five types of uncertainties and their implications for likelihood of adoption, in terms of feasibility, desirability and usefulness from the researchers' perspectives.

FINDINGS

A predominant theme among interviewees' accounts of their research was the overwhelming need to manage uncertainties at every stage of the research process. In this section, we discuss five different types of uncertainties faced by researchers related to process, data, knowledge, delegation and quality.

I. Process Uncertainties

> The public generally thinks that scientists work like Sherlock Holmes, but really most of us work like Peter Whimsy...it's nonlinear in the sense that you don't know where things are going to lead. [S4, paleontologist]

> When you're at the cutting edge, you don't know for sure. [S10, astronomer]

These quotes succinctly capture a pervasive challenge described by all our interviewees: the process of discovery can be highly uncertain, iterative, and often serendipitous. This theme was further echoed when researchers found themselves crafting a story to explain how they arrived at findings, even though it often was not "how it happened." The observation that research is serendipitous and iterative is certainly not new: several prior studies [27, 48, 55] showed that scientists use the *unexpected* as a way to generate new experiments and theories. Our interview data suggests that this trend is a pervasive feature of research across disciplines, including the humanities.

For many researchers, the ill-defined nature of a research agenda's trajectory can make it hard to imagine how crowdsourcing might play a role, specifically when and how research tasks could or should be delegated to crowdworkers or volunteers. Some researchers were reluctant to delegate tasks until the research question became clear and the tasks could become "mechanical" (*e.g.*, having students dig up specific pieces of data or information to support the claims of the research [S7, sociologist]), which typically did not happen until the final stages of the research process. Their reluctance to delegate reflected the need to control the uncertainties of the whole process, and suggests a likely mismatch to crowdsourcing techniques. For example, tasks whose outcome could have substantive downstream impacts were rarely delegated, for fear that the person handling the task might lack the *expertise* to "recognize the alternative ways of doing things, and the implications of that for the next stages," or adequate *curiosity* to explore the space of possibilities [S7, sociologist].

A second observation was that *serendipity*, an important factor in discovery, could come from two different sources. Serendipity could be exogenous, surfaced through chance encounters with people who hold critical pieces of information needed to solve the problem or people who bring diverse perspectives and backgrounds. It could also be endogenous, originating from the researchers' own close examinations of raw data, which is common in qualitative research, *e.g.*, where transcription may be considered a part of the analysis process.

In the former case, crowdsourcing could become a useful tool, as it can support many individuals inspecting the same data and potentially contributing novel insights. Some of the existing citizen science platforms are already designed to facilitate serendipitous discoveries. In a well-known example from Galaxy Zoo [51], a small group of volunteers identified what are now known as the "Green Pea" galaxies [11]; this discovery has led to new research [44, 42, 14] that may have otherwise not taken place. As an astrophysicist [S3] put it, "I thought it was wonderful that a non-astronomer found [the Hanny's *Voorwerp*]. And then it turns out there are other objects like that around if you know what to look for." Using such human insight, researchers could then "design algorithms to look for [what] no one was even looking for." A paleontologist [S4] recalled a situation where he showed images of microfossils at a talk and someone in the audience, to his pleasant surprise, identified them: "we are all victims of our own experience. There always is the possibility that the right answer lies out there just beyond your expertise." This example points to the limits of expertise and benefits of connecting researchers to not only non-experts, who in large numbers can help pick out the needle in the haystack, but also to experts, who may hold a key piece of missing knowledge for the problem at hand.

In the case where new insights originate in the expert's own close examinations and interpretations of data, crowdsourcing may seem less desirable. Some researchers argued that processing data, though laborious and time-consuming, was "an act of thinking," and outsourcing such data processing tasks (*e.g.*, to a crowdsourcing platform) may diminish the whole "joy of discovery". As an historian described:

> When I talk to people in other fields, they can't believe I do this, but I literally transcribe all this material, line after line, so that when you get to your thousandth object, you start to see patterns and trends you had never known before... The real reason is—it's simply the act of ploughing through it. It's in the act of yourself digging through it that you see a matrix that you wouldn't see if you just looked at the individual finding. So you actually have to see the data in its original context to begin to understand what that matrix is. [S18, historian]

Or, as a neurologist told us,

> If somebody else tells you "hey, there's something really interesting in your data set," it may take some of the fun out of it. [S8, neurologist]

For researchers who worked closely with raw data (*e.g.*, [S18, historian], [S7, sociologist], [S8, neurologist]), the process of analysis was inseparable from the process of collecting and transforming data. These researchers expressed a need for iterative work by the same individual to develop new insights. For example, a sociologist [S7] described the initial stage of research as being "very non-linear". He remarked: "Your focus might change, and your findings might be surprising given what you expect. You try to make sense of them and find out what it means, you start considering other factors that you didn't initially think were important." While collecting data, a historian [S17] noticed "a completely anomalous set of information [that] actually seemed to fit a pattern ...which contradicted everything that we thought we knew." He reacted by "multiplying the evidence, looking in new places," and eventually made an unexpected discovery.

These uncertainties in the *non-linear* and *serendipitous* research processes had several impacts on the feasibility and desirability of crowdsourcing as a research tool. Researchers may not want to incur upfront cost of setting up a crowdsourcing project if they foresee their research questions changing. In some cases, crowdsourcing may not even be feasible until the later stages of the research process, when the research questions solidify after several iterations over the data, the process for answering the questions is concrete and can be clearly articulated, and broken down into a well-defined workflow to create delegatable tasks. Serendipity relies upon either outside perspectives or close engagement with the research data in order to surface new insights. We posit that when external perspectives can provide new value, crowdsourcing may be deemed a suitable approach, while processes that rely on repeated or comprehensive examination of data by a single individual appear to be less amenable to crowd participation.

II. Data Uncertainties

Crowdsourcing often requires publicly revealing raw data and research questions, which are then subject to scrutiny and peer evaluation, creating uncertainties related to the *data*.

Sharing raw data can sometimes be problematic due to its sensitive nature; this can rule out crowdsourcing as a potential solution unless safeguards are in place, *e.g.*, allowing researchers to keep their project "private" and shared with only a small group of close colleagues. For example, patient data [S8, neurologist], lab animal data [S5, neurologist], and student data [S14, psychologist] were never shared due to strict ethics regulations. The whereabouts of certain endangered plants [29] and archaeological sites [S15, archaeologist] were not shared due to potential for looting or destruction.

Second, we observed a relationship between researchers' perceptions of the abundance of research questions the data could support and their openness to sharing data publicly. This relationship hints at a potential risk of crowdsourcing, which is the exposure of research questions that researchers consider to be scarce or rare. As one interviewee put it,

> There are two worlds. One world where there are too many good ideas, and we just don't have time and resources to pursue them. That's a world of abundance. The other is a world of scarcity, where there aren't enough good ideas, and we need to hoard and protect them. [S14, psychologist]

In some projects, it seemed that the number of research questions to ask about a data set was endless and the chance of two questions being exactly the same was slim, or that interesting questions could only be asked of adequately large quantities of data. A political scientist said,

> Astronomers have the same problem as we do. You can see things much more clearly if you do a big data collection than if you do a lot of uncoordinated activities—you

> don't get as big a telescope and you don't have the telescopes all pointed where they need to be. [S1, political scientist]

Similarly in ecology,

> You don't need to protect the data. Collecting the data is only the beginning. When we've got as much data as there are there, analysis becomes kind of a challenge. ... There's almost no end of questions that can be asked or hypotheses that can be tested. [S6, ecologist]

Where projects were considered *question-rich*, exposing the raw data (*e.g.*, through crowdsourcing) would not be a strong concern. In contrast, when data were sparse and difficult or expensive to obtain (*e.g.*, new specimens from a remote region of the world [S9, biologist]), the rarity of the data and the high cost of obtaining it warranted caution. In such situations, researchers were inclined to expose data only *after* publication, as the quotes below illustrate, and as previously reported in studies of data sharing practices [80].

> There's a big imperative of course from the NSF to put your data online after you've finished your study. ... It's still a very sensitive issue in my field. People are not crazy about doing this because ... they don't want to be scooped ... before they're done, [and] sometimes they take forever to finish their data. They don't want somebody to come along and publish before they do. [S9, biologist]

> My aspiration is that [sharing data] will become customary among historians ... once we've published it. You've got to get the rewards of your labor, but once you've done that, it only makes sense to [share]. Data collection is expensive because ... you have to create it from scratch. [S17, historian]

In summary, data uncertainties reflected researchers' sense of abundance or scarcity in both data and research questions. In situations of abundance (*e.g.*, political science, astronomy), crowdsourcing has been embraced as a feasible approach to address some research questions. In situations of perceived scarcity, most researchers were likely to either shy away from crowdsourcing that requires making their data public, or found crowdsourcing undesirable because their processing needs did not seem to warrant the effort and up front cost of setting up a project. Finally, researchers' perceptions of abundance versus scarcity and the extent of their concerns over peer inspection depended not only on the particular project, but also on the sharing norms of their respective disciplines. This suggests that crowdsourcing may be more acceptable in fields with well established data sharing practices.

III. Knowledge Uncertainties

Expertise is a complex concept that has been rigorously redefined and re-examined [17]; here, we use the term expertise loosely to mean both explicit knowledge [26] and implicit (or tacit) knowledge [61]. Explicit knowledge is concrete information that can be taught to novices through detailed instructions. Implicit knowledge is often represented as know-how that is acquired through immersion and guided practice. The feasibility of crowdsourcing can depend on which type of knowledge the tasks require. Tasks may be more amenable to crowdsourcing if they require minimal knowledge that can be communicated concisely, rather than extensive knowledge that requires years of training. The latter may require the apprenticeship model of training, such as direct one-on-one mentoring or working with a small group of peers, which is a predominant mode of knowledge transfer in current research practice and more difficult to scale to crowdsourcing contexts.

One class of tasks that required extensive domain knowledge were tasks with *ambiguity*, such as when data were noisy (*e.g.*, a blurry image), or when the rules (*e.g.*, for classifying an image) were ill-defined, leaving details open for interpretation. Researchers are trained to handle such ambiguity throughout the entire research pipeline, from collection and transformation to analysis; interviewees discussed several types of ambiguity that could impose operational challenges for crowdsourcing. A neurologist gave an example of ambiguity that requires training to manage effectively:

> When a behavior is well-established, [*e.g.*,] if an animal is up and moving around, nobody will disagree on that. That's easy. But when a mouse first wakes up ... they don't show a lot of facial expressions, and they don't get very animated ... the brain waves might show that they're awake but maybe it kind of looks a little sleepy, and they're just sitting there, zoning out. How are you going to call that? This is where the subjectivity comes in. [S8, neurologist]

An archaeologist [S15] similarly identified multiple points of ambiguity for a core research task:

> How highly mounded does a site have to be? Does it need to be a meter, fifty centimeters, or a ten centimeter rise? Is that too ephemeral to be considered a site? And then you can get into questions of artifact density. How dense do they have to be? Do you want to measure it? Do you want to talk about artifacts per square meter? [S15, archaeologist]

As this quote suggests, when the task was ambiguous, extensive domain knowledge was often needed. Sorting out the idiosyncratic nature of the data—determining what is relevant versus irrelevant given the specific research questions, making the correct inferences, separating "glitches" from results that are in fact valid but unusual [S2, systems biologist], and "stretching the bounds of what's credible" [S6, ecologist]—were all believed to require extensive expertise.

> If you have that sequence, what could it possibly be? Well, it could be a U and an M if you connected it this way. Or it could be an IMI if you connected it this way... That's where you have to know what's feasible in terms of Latin... I know that in a document which involves a peasant, it's very likely that that sequence is going to be a *vinum* because this is a wine-producing area [where] they have a lot of vineyards. [S18, historian]

This quote identifies the importance of domain knowledge for completing a transcription task. A biologist explained a similar need for procedural and technical knowledge to interpret data:

> You look at the data and say "wait, this isn't something I would've expected!" ... You really need to have some level of technical expertise to be able to separate between that artifact is caused by experiment and that artifact is caused by biology. [S2, systems biologist]

Many researchers also characterized the expertise required in their research work as something that could only be accumulated through experience and apprenticeship, and difficult to teach in a short period of time. For example, an historian [S18] said that transcription was not like a science, but more akin to "riding a bicycle" where "you just have to do it and get a feel for it." This particular researcher adopted the strategy of organizing a weekly "paleography slam" workshop, where students would bring documents and ask others for ideas on how to transcribe certain words. As a result, "everyone gets better ... and gets a sense of the landscape of possibilities by looking at how it's done here." He remarked, "it's almost a crowdsourcing idea, but obviously it is in a smaller group."

Similarly, a systems biologist [S2] referred to lab work as an *artisanal* skill, like "training someone to be a baker" with a "prescribed set of techniques that have to be learned through experiment [that is] not easy to completely outsource." This type of experience-based knowledge was teachable, but required a senior researcher to closely monitor and guide the work on an individual basis.

> There's lots and lots of little traps ... It's not like there's a book out there that says these are the pitfalls for making record linkages and the transcription errors that you are capable of making. It's almost entirely done by an apprentice-like system. [S18, historian]

Crowdsourcing requires some level of *delegation* and giving up control of a task that was traditionally done by the researcher herself or a small team of students and colleagues. For current research practices, delegation was a sensitive, cost-benefit decision [36], that added to the already overwhelming uncertainties of the research process. A psychologist explicitly demonstrated how such cost-benefit considerations might affect decisions about whether to use crowdsourcing:

> Are you throwing out good ideas by doing crappy tests of them using fast, cheap Mechanical Turk? That's a tough trade-off. There's this constant calculation of an expected value: probability that it's real, probability of robustness, cost of collecting and completing the project. All of that, multiplied by the likelihood that it's an important and interesting project. [S14, psychologist]

Delegation also incurred a cost in terms of time. It can be non-trivial to match the work to an individual with the right skills, attitude, and interest, and to train and monitor them. It also takes time to deal with the mistakes that non-experts make. In the worst case, mistakes may be irreversible, where there is only *one shot* of getting it right, *e.g.*, when there was a limited amount of data [S4, paleontologist], or when the data collection procedure (*e.g.*, preparation of an animal for experimentation) was costly in terms of time [S5 neurologist; S8, neurologist]. For example, a neurologist [S5] said,

> Scoring [data] is actually easier, because if they do it wrong, we can do it again. Gathering up the data for the deletion mice—we have one shot. If something is screwed up, like the camera is in the wrong position, those measurements are not worth anything anymore. [S5, neurologist]

Even if the mistakes were reversible, they could take more time to correct than to do the work in the first place. As a systems biologist [S2] said, "There's always this balance ... cheap labor is great but [not] if it actually takes more time to manage the labor." While this is common consideration for any academic apprenticeship, the unknown human resource requirements for supporting the efforts of crowdworkers complicates the cost-benefit equation and creates an additional source of uncertainty.

In summary, knowledge uncertainties could be predicted by the degree of explicit versus implicit knowledge required for the work. Where the tasks are explicit requiring minimal, easily communicable knowledge, crowdsourcing would be a more feasible option. When the tasks require extensive and obscure knowledge or implicit know-how, innovative solutions (*e.g.*, in terms of training, decomposition, etc) are required to effectively delegate to an inexperienced crowd.

IV. Delegation Uncertainties

Whether or not researchers were comfortable with delegation hinged not only on their perceptions of the research process and the crowd's abilities to perform knowledge-intensive tasks (as previously discussed) or carry out well-grounded interpretation of artifacts, but also on the researcher's perceptions of the public's intentions and interests. Prior interactions with the public around science communication could influence researchers' decision to avoid or embrace crowdsourcing. Understanding these perceptions helped clarify researchers' concerns about engaging crowdworkers: positive interactions highlighted the potential benefits of working with the crowd, while negative interactions underscored the challenges. Several of our interviewees [S12, biologist; S13, physicist; S4, paleontologist; S1, political scientist] regularly interacted with the general public through talks and correspondence. Their experiences with these encounters offered a starting point to understand their perceptions of public contributors.

A few of the researchers mentioned being regularly approached by people from the general public who were fascinated by science. A political scientist [S1] recounted a time when "one older couple wrote ... an email and asked how [they] would go about assessing election performance, ... [then] assembled a database, and ... did an analysis." Here, seeing capability and interest, the researcher assumed the role of a colleague [76] and facilitated their independent inquiry. This scenario is far from unique [70].

In contrast, other researchers were skeptical that anyone outside of their scientific community would be interested in looking at their data, and therefore may not trust that a crowdsourcing project would attract adequate participation. As some researchers opined, "there's no way the public would ask to locate the medulla of a mouse," [S5, neurologist] or would want to “record electroencephalogram patterns on the weekend” [S8, neurologist].

Even those who understood the enthusiasm of volunteers were wary of the fact that the quality of crowdsourced data may be “spotty” at best, with a few gems and a lot of erroneous and naive interpretations. One of our researchers said,

> In the early days of the ... mission to Mars, when all the images became available to the public as they’re taken, ...we would get tons of emails. Some of them were geologists who had actually some very interesting things to say. [But then] there was one guy who kept writing to see why was it that we couldn’t see that this outcrop was a pleiosaur skeleton... In a sense, I think it is the “yin” of citizen science because there are such varying experiences, you’re going to get observations or interpretations of the images that go from the sublime to the ridiculous. [S4, paleontologist]

Or, as an archaeologist noted,

> There is a cottage industry of avocational archaeologists who just scour the world on Google Earth, looking at the images ... in some cases making really significant discoveries ... but in a lot of cases, the interpretations are extraordinarily naive. [S15, archeologist]

Lastly, some researchers mentioned fear of *sabotage* if the data were open to examination. Researchers expressed consistent concern that competitors and those who disagreed with the research methodologies (*e.g.*, animal activists) may intentionally produce bad data if participation was open to all rather than constrained to known parties.

> One issue with sharing animal research data is there are some people who really disapprove of any animal research, even in mice... What if somebody says “oh sure, I’ll score that data,” and then they intentionally do it in a bad way. That would be really, really disruptive to have a saboteur in the midst. [S8, neurologist]

Several researchers also mentioned the importance of a *human connection* in choosing which students they trained and worked with. This sensibility and desire for connectedness to collaborators, for either its social benefits or research quality impacts, may also influence researchers’ willingness to engage with self-selected, transient and semi-anonymous crowdworkers instead of hand-picked protégés. As an historian [S17] remarked, “I choose very carefully who the crowd is ... We have a human relationship. I have some knowledge of their personality and their quality of mind. There’s some human link, some individual link, and I find that quite important.” He explained that the quality of the results could depend on appropriately matching questions to data, and data to students.

Finally, researchers faced *role uncertainties*—whether and what to delegate sometimes presented a moral dilemma. On the one hand, researchers relied on students to get research tasks done. At the same time, their duty was to train the next generation of scholars. Simple, repetitive tasks could have educational value in some cases or become “drudgery” in others. The academic apprenticeship is considered a *quid pro quo* (if sometimes uneven) exchange of mentorship for labor, but without the explicit promise of training and mentorship, asking the crowd to undertake unattractive tasks could be perceived as exploitative. Therefore, whether or not to delegate a particular task can become an ethical issue. For instance, an astrophysicist gave examples of both perspectives on whether to ask humans to do tasks that machines could complete:

> This guy ... had an army of undergrads, a dozen at a time, and their task was to extract a spectral feature—take an image, take a finding chart and click on the known stars... I was like, “This is drudgery! I could automate these tasks and you could fire all these undergrads.” And he would say, “No, that’s not the point. They need something to do.” ... But I just can’t do that. I can’t give a student something I can trivially automate. [S3, astrophysicist]

The potential for exploitation is an inherent ethical dilemma in research-oriented crowdsourcing systems. Prior work raised concerns about the use of crowdworkers for research on paid [53, 30] and volunteer-based crowdsourcing platforms [68]. Is crowdsourcing voluntary contributions to research a form of paid labor, a self-interested pursuit, or a way to educate the public? These *role uncertainties* pose another challenge for tailoring crowdsourcing systems to support multiple goals.

On many existing crowdsourcing platforms, workers are “faceless” and transient, *i.e.*, they come and go without ever developing working relationships. The cloak of invisibility around their unknown characteristics — abilities, intentions, interests — further adds to the sense of uncertainty, making the legitimacy of delegation questionable. This scenario is somewhat less true in certain citizen science projects where crowds are more realistically “semi-anonymous” and can be involved in long-term participation. Nonetheless, the lack of direct interaction can be unsettling for researchers new to this paradigm. Concerns around knowledge and delegation uncertainties hinted at the enduring duty of researchers as educators, and may contribute to some discomfort with crowdsourcing.

V. Quality Uncertainties

Many researchers expressed a feeling of apprehension and vulnerability about exposing research data that they produced. As one interviewee said,

> If every time we published an experiment, we basically [shared] our data in a way that anybody could look at it, we would have to feel very confident that what we are putting out there is really, really good. [S8, neurologist]

This sense of vulnerability, which already exists in current research practices, is problematic in that it can only be exacerbated in the crowdsourcing contexts where researchers have even less control. Our interview data identified several ways researchers reacted to and addressed quality uncertainties.

First, several researchers mentioned that they preferred delegating to *machine computation* rather than humans. The strength of and rationale behind this preference revealed perceptions and beliefs about quality. For example, one researcher said,

> If our algorithms can't make a decision, then it's a poorly defined question. We have mathematical things that are defined and we can compute them with a computer exactly and we don't need any humans involved. [S3, astrophysicist]

This researcher used automation as a forcing function to ensure that the research questions were crisply and unambiguously defined. Since all the difficult decisions were taken out of the process, delegation to automation was feasible and likely to produce high quality results.

In other cases, researchers used automation to generate data to steer their inquiries. For example, an archaeologist [S15] used satellite imagery to teach a machine to recognize properties of archaeological sites (*e.g.*, topography, soil discoloration, surface artifacts), and would then visit the most promising sites. Likewise, a behavioral economist [S11] used topic modeling to analyze open-ended questions, instead of relying on manual coding. A sociologist [S7] contemplated enumerating and computing the best combination of a large set of variables to find the best model, instead of the traditional trial and error approach. In all these scenarios, automation produced abundant, cheap and potentially lower quality work which served as a source of inspiration for the expert.

While some researchers may be reluctant to rely on human processing, in practice, automation and crowdsourcing provide a common benefit, namely the ability to amass large quantity of data quickly in order to generate quick insight. As [S8, neurologist] said, "[during busy times], if there was a way to have it out there...where some info trickled back on it, that potentially could influence how eagerly [a researcher] would then pursue the next bit."

In fact, exclusive delegation to automation is rare. Many researchers instead employed *semi-automated* techniques, with the computer making guesses and experts double-checking the answers or deciphering edge cases. A neurologist said,

> If you have a lot of different type of gait abnormalities, you'd be spending all your time rearranging the system to do what it's supposed to do. So what we are doing now is we just do it kind of semi-manually — we have a Matlab program where we download the video data into it and we [manually] determine when the foot hits the plexiglass. [S5, neurologist]

This researcher had a reasonably well-established process, but one in which unexpected situations often arose. He used machine computation to automate the tedious parts, and dealt with quality issues by leaving the critical interpretation of the data to the experts; a parallel strategy with human computation is seen in many citizen science projects.

These observations pointed to the fact that data quality was not an isolated concept, but was often considered alongside with quantity (*i.e.*, how much data can be gathered or produced), speed (*i.e.*, how quickly can data be gathered or produced) and the convergence of evidence (*i.e.*, whether the data contribute an additional, orthogonal source of evidence to confirm/disconfirm the hypothesis in question). The consideration of crowdsourcing as a research tool may surface similar kinds of cost-benefit tradeoffs, so we asked researchers whether they would have any use for large amounts of lower quality data. They agreed that the level of quality needed would often depend on the granularity of the research questions: the coarser the question, the more likely annotators will agree, and the less demand for precision in each individual's answer.

> In some experiments, a sort of bird's eye view is sufficient—if the question is "does this drug increase sleep?", then all we care about is the amount of sleep. Even an undergrad's assessment of that data set would probably be sufficient to answer that yes-or-no question which would then dictate whether or not we say "yes, let's take that data and then analyze it in more detail." [In contrast, for the question] "Exactly how good are these mice at staying awake for periods of forty minutes or more?" There, even one little lapse into sleep somewhere during that forty minute interval is important, so the data needs to be gone over extremely carefully. [S8, neurologist]

An archaeologist identified similar interdependencies between research questions and task performance that could impact the utility of crowdsourcing:

> I would rather have a broad area with sites that I've defined subjectively, with a relatively high amount of error, than a small area where I'm absolutely certain; that stems from the nature of my questions. Urbanism and these settlement patterns—these are regional phenomena. If I know a little tiny corner of an agricultural plain really well, it's quite likely that I'm missing out on archaeological sites elsewhere that would really alter my interpretation. [S15, archaeologist]

Several researchers also expressed a preference for speed in lieu of impeccable data. Similar to other aspects of crowdsourcing, the speed at which work might be done was an uncertainty—but likely perceived as an advantage. An historian [S17] noted that technology lets researchers "rake through huge amounts of sources very cheaply... You could ask stupid questions but it didn't matter because you failed in about two seconds. So you could afford to fail again and again."

Finally, lower quality data may be useful if they provide multiple sources of evidence to address the same research question. Although crowdsourcing is sometimes shown to generate data equivalent to professionals, the usual expectation is that crowdworkers will generate somewhat inferior data [10]. Researchers mentioned *triangulation*—collecting multiple types of evidence, or measuring the same thing with multiple instruments, to arrive at the same conclusions—as a familiar strategy for ensuring reliability often expected by reviewers.

> Someone once called it "two-versality" — you show it once, who knows if it's true? If you show it twice, even if it's two crappy ways of showing it, it must be true, right? If you find two cases of something it must be a universal process. And so that's a standard thing. If you show something by any technique, people want to see

> you show something by another technique, to validate something. [S2, systems biologist]

> There are four or five of these [tests], all of which are going to reflect water column oxidation, either locally or more broadly. All of them have potential traps. But since they're different traps, if they all agree, then your ... confidence is increased by having multiple techniques give you the same answer. [S4, paleontologist]

While crowdsourcing has the potential to serve as triangulation for other methods, the more likely expectation is that crowdsourced data may require additional triangulation for confidence in results, which could increase the research workload and the complexity of the design dependencies.

Implicit in some of the responses was the assumption that machine computation is "objective" while human work is not, and more importantly, that machine computation is cheap, fast and scalable. Crowdsourcing may be undesirable to researchers who require absolute precision, but may be welcomed by those with different quality-convenience tradeoffs in mind, such as the desire to collect massive amounts of data, collect data quickly, or collect diverse sources of evidence to support the same conclusion. At the same time, crowdsourcing may incur additional effort for verification if the quality of the work is below standard expectations for the field.

DISCUSSION

Under the right conditions, crowdsourcing can be a valuable tool for streamlining research processes. As mentioned earlier, four of our interviewees already employed crowdsourcing to some extent: running pilot surveys on Mechanical Turk [S14, psychologist; S1, political scientist], leading a SETI@home-like project where contributors donated computer processing time to help search for molecules [S16, chemist], and engaging students to help with mapping shipwrecks [S17, historian]. However, the uncertainties that researchers faced—around process, data, knowledge, delegation, and quality—made it difficult to adequately assess whether crowdsourcing would be a feasible, desirable, or useful tool for their scholarship. Here, we summarize the key findings to these research questions and discuss their implications for design, research, and practice.

Challenges: Feasibility, Desirability and Usefulness

The reasons why crowdsourcing could be mis-aligned with the needs of researchers clustered into three types of concerns—feasibility (i.e., that it may be impossible to adopt crowdsourcing in the first place), desirability (i.e., that researchers were not comfortable enough with adopting a new approach like crowdsourcing) and utility (i.e., that researchers might not recognize the benefits of crowdsourcing as a research tool).

Our interview data suggested that crowdsourcing may be less feasible for a variety of reasons: if the workflow is not decomposable, if data processing and analysis are intertwined rather than sequentially connected, if serendipity depends on deep knowledge of the raw material or extensive subject matter knowledge that only an expert researcher can accrue, and if artifacts and data contain sensitive, private information that cannot be shared. In any of these cases, adoption of crowdsourcing may offer little advantage for researchers.

Crowdsourcing may not be deemed desirable by researchers who experience the research process as inherently iterative and non-linear, with research questions remaining open-ended until the end stages of the process. Furthermore, if the research data are scarce, difficult or costly to obtain, or perceived to support only a few research questions, researchers may be reluctant to expose them. Such protective attitudes were more evident in disciplines where open data sharing norms are not yet well established. Crowdsourcing can be desirable, however, when it achieves the right combination of production quality, speed, and quantity for a particular project. This balance often depends on the granularity of the research questions and how researchers think about quality-convenience tradeoffs. Similar to machine computation, the low-quality high-throughput data that crowdsourcing generates can potentially provide significant utility to researchers, by generating rough insights to facilitate serendipitous discoveries and by handling the tedious tasks, leaving the important work to human experts.

Researchers may also doubt the utility of crowdsourcing due to (often unsubstantiated) perceptions of the crowd's abilities, intentions, and interests, any of which could have a direct impact on research quality and peer evaluations of the legitimacy of the crowdsourcing project. They may also feel conflicted about delegating potentially tedious tasks to contributors whom they may feel compelled to educate and inspire. The usefulness of crowdsourcing can also be limited when tasks require extensive specialized knowledge or implicit know-how typically obtained through experience, since there are significant costs to verifying the quality of the results or repairing errors.

Existing crowdsourcing platforms address these problems of feasibility, desirability and usefulness, but only partially. Our work surfaces these issues as grounds for future studies and tool development aimed at reducing the mismatch between research and crowdsourcing.

Implications

To realize the potential of crowdsourcing as a research tool despite the uncertainties faced by researchers, we need to design better features for researcher-centric crowdsourcing platforms and provide reliable decision-making guidelines and tools for practitioners and policy makers. Below, we discuss the implications of our work for platform and guideline design, as well as outline open questions.

Desiderata for Researcher-Centric Crowdsourcing Platforms

Existing easy-to-use crowdsourcing tools are particularly well-suited to research projects with stable goals, decomposable workflows, limited concerns about data sharing, low expertise requirement, etc. Table 3 provides some well-known examples of crowdsourcing projects with knowledge production goals, whose characteristics make crowdsourcing a feasible, desirable and useful tool. We also include an example of a general class of projects, water quality monitoring, for which crowdsourcing has variable value, due in part to the tendency of these projects to have a highly localized focus and purpose. Similarly, the example of Casey Trees [12] is a project that

	Galaxy Zoo	Foldit	eBird	Ancient Lives	Water Quality	Casey Trees
Process Uncertainties						
research goals	established	established	established	established	variable	variable
workflow	decomposable	decomposable	semi-decomposed	decomposable	decomposable	decomposable
processing & analysis	separable	separable	semi-separable	separable	separable	separable
serendipity	diversity	diversity	diversity, volume	diversity	N/A	N/A
Data Uncertainties						
sensitivity	shareable	shareable	mostly shared	private	variable	shareable
typicality	common	N/A	N/A	rare	variable	localized
quantity	abundant	N/A	N/A	abundant	limited	limited
availability	limited	N/A	limited, diffuse	limited	very limited	very limited
research questions	many	many	many	many	highly specific	several
sharing norms	open	open	open	closed	variable	N/A
Knowledge Uncertainties						
requirements	explicit	semi-explicit	explicit & implicit	explicit	semi-explicit	explicit
availability	common	common	common	common	uncommon	uncommon
Delegation Uncertainties						
crowd interest	interested	interested	rabid	interested	localized	localized
crowd intention	good	good	good	good	self-interested	good
crowd ability	capable	range	range	capable	range	range
role of researcher	manager	manager	data user	data user	analyst	outreach
Quality Uncertainties						
tolerance for ambiguity	high	low	high	low	low	medium
goals	quantity	precision, diversity	quantity, coverage	quantity	regulatory precision	coverage

Table 3. Example of several well-known projects according to the criteria for feasibility, desirability and usefulness of crowdsourcing.

collects data to inform urban forestry efforts in the Washington DC area; despite a very large local population base, the nature of involvement means that the project bears more resemblance to small team collaboration than crowdsourcing.

These examples make it apparent that projects that have *all* of these attributes (stable goals, decomposable workflows, and so forth) are rare. Changes in any of these attribute values can mean either higher up front costs and/or lower benefits of employing a crowd-powered solution. The technical agenda for designing effective researcher-centric crowdsourcing platform seems clear: we need to develop solutions that lower the cost and increase the benefit for projects with attributes that do not match current crowd tools.

One example attribute where many research projects do not match the assumption of today's crowdsourcing tools is the stability of the research goals. Most research projects follow an iterative process before settling on a clear research question and established process where scaling up would be valuable. One possibility is to structure crowdsourcing platforms to provide much more control to the researchers, *e.g.*, to use the platform at first as a personal tool for oneself or a small team of trusted colleagues, with the ability to fluidly switch between close examination of the raw data versus reaching out to the "crowd" for additional insights or help with specific sub-tasks. By providing the ability to gradually publicize the projects—first to closest collaborators, then to fellow experts, then to broader crowds—researchers could better handle the iterative nature of the research process. Early on, the tasks are coarse and ill-defined, so only experts can contribute; when the real goal or process is found, a more carefully decomposed workflow that can robustly accommodate novices can be put in place.

Even though this type of scaffolding may already exist in current citizen science project management practices, better technological solutions, *e.g.*, platform features that facilitate small-team collaboration in addition to large-scale crowdsourcing, would help streamline the process. This suggests creating a new kind of one-size-fits-most research-oriented crowdsourcing platform that serves both communities and crowds, or alternatively, developing suites of specialized tools to support diverse crowdsourcing projects or deployment of crowdsourcing at different stages of the research process.

To lower the barrier to launching a crowdsourcing project, future platforms should provide more effective end-user tools for designing, monitoring and re-designing workflows. Although several tools already exist (*e.g.*, CrowdCurio [50], Panoptes [9], CrowdCrafting [71]), there has been little work on understanding what makes these authoring interfaces effective or usable from the researchers' perspective.

Another fruitful area for further research and development would be new ways for researchers to accurately and quickly assess impacts of uncertainties around process, data, knowledge, delegation and quality, to help make sound decisions about the cost-benefit tradeoffs of using crowdsourcing for a particular project. As an example, visualizations that enable researchers to distinguish between well-intentioned versus malicious, capable versus incapable, and interested versus uninterested contributors, could reduce uncertainties around delegation. Better mechanisms for researchers to achieve greater control over the type of "crowd" being engaged at different stages in the research may also help mitigate concerns. While some systems currently require volunteers to pre-qualify with a quiz (*e.g.*, Stardust@Home [86]), the typical binary go/no-go style of filtering participants wastes much potential and can damage good will. Given the known breadth of performance and diverse abilities of the crowd, crafting a portfolio of participation options (*e.g.*, following the reader-to-leader model [62]) may offer a stronger mechanism for directing participants to the tasks that they can perform best. This would require system features to support profiling participants' skills, robust task routing, and managing process interdependencies.

As another example, what tools might help researchers make well-founded decisions about the tradeoffs between quality, quantity, speed and diversity of data sources? Our interview data showed that quality is a complex, multi-faceted consideration, where the necessary level of quality can depend on the actual purpose of the crowdsourcing and other quality-convenience tradeoffs. For most existing crowdsourcing platforms, redundancy (or consensus) is the predominant method for ensuring correctness—if multiple independent workers do the same task and arrive at the same answer, we consider the answer reliable. Our interview data suggests that current academic research practices prefer to use triangulation to signal research quality, and that replication [16] is surprisingly uncommon due to the lack of funding, the repetitive and labor-intensive nature of the task, and concerns around resolving disagreement between multiple evaluators. For example, in human sleep studies, a single researcher would typically manually annotate hundreds of thousands of EEG epochs [S8, neurologist]. To have multiple student annotators process the same EEG records would be "tripling the work" and considered "cruel and unusual punishment." With these challenges in mind, it may be useful to develop tools or processes that help researchers assess the level of quality and participation needed from crowdsourcing in order to test specific hypotheses or achieve adequate coverage, density, or replication to answer research questions.

Given the fast growth of research based on crowdsourcing and its multidisciplinary nature, creating a comprehensive reference library of descriptive metadata about crowdsourcing projects could enable analyses to support new kinds of platform features for addressing the challenges of assessing different types of uncertainties. By enabling researchers to more easily identify prior work that involved similar parameters to their own ideas, they could better weigh potential risks against expected productivity and adopt tools that are most appropriate for their specific needs.

Guidelines for Evaluating Crowdsourced Projects

Decisions about crowdsourcing involve both peers (grant and paper reviewers) and those who make and implement policies that impact science practice (*e.g.*, funding program officers). The results of this study can help decision makers identify those projects for which existing crowdsourcing methods are already adequate and little additional justification is needed for including crowdsourcing in the proposal. For other projects, these findings can help researchers anticipate potential challenges and direct their effort toward demonstrating how these challenges will be overcome. Likewise, our work provides a "checklist" of dimensions (*i.e.*, Table 1) that a proposer needs to address with care.

Another area for research in service to practice and policy is the development of instruments and software for evaluating implementations of research-oriented crowdsourcing tools. While instruments have been developed for participant outcomes evaluation in citizen science (*e.g.*, the DEVISE scales[1]), there are no comparable or comprehensive practitioner-oriented tools for evaluating the costs to contributors, the research resource requirements and return on investment ([22] and [43] provide initial metrics), or the adherence to ethical frameworks such as the Ethical Principles for Participatory Research Design [8]. Work supporting development of automatic detection of problematic changes in community social dynamics, for example, could allow practitioners to continually monitor and assess the health of their crowdsourcing projects, enabling timely interventions to eliminate sources of confusion or contention and to retain experienced, high-value contributors.

Open Questions for CSCW Researchers

Finally, our work opens up a set of new research questions, several of which are particularly well suited to CSCW research. For example:

- How do peer perceptions of crowdsourcing empirically vary by disciplines and disciplinary features, *e.g.*, as related to open data norms?
- How can empirical analysis of existing research-oriented crowdsourcing projects refine the taxonomy of uncertainties identified in this study?
- What techniques can provide useful and scalable ways to match individual participants to tasks, moving beyond binary determinations of ability to more nuanced and functional characterization of contributors?
- What conditions determine whether or not payment is appropriate or necessary to incentivize participation?
- If researchers are able to dynamically shift the composition of the crowd for their projects, how does the evolving organizational structure, from small teams to medium-sized community to large anonymous crowds, influence the outcomes in terms of research results and participant benefit?

Recent research in crowdsourcing has introduced a variety of new interaction paradigms that enable non-experts to learn and perform complex, expert-level tasks [57, 67, 79]. The ideas and lessons learned have not yet made their way into citizen science or research-oriented crowdsourcing, suggesting many opportunities for future work. Likewise, applications of cutting edge research on technical and policy solutions for sharing sensitive research data in a privacy preserving manner (*e.g.*, [40]) merit exploration. Future worker could focus on equivalent solutions for researcher-centric crowdsourcing.

CONCLUSION

Attending to the social norms, practices and values of researchers is critical to designing successful research-oriented crowdsourcing technologies. Our paper contributes a formative study of the needs and constraints of researchers as potential users of crowdsourcing systems. In this work, we answered two research questions—Under what circumstances is crowdsourcing a feasible, desirable or useful tool for researchers? Under what circumstances is crowdsourcing not suitable for research?—by identifying a set of uncertainties related to research process, data, knowledge, delegation and quality, as well as the conditions making the research more or less amenable to crowdsourcing. This paper contributes to the CSCW knowledge base on large-scale collaboration by delineating the types of uncertainties, across a wide range of disciplines and research contexts, that can help inform the design of future researcher-centric crowdsourcing systems and projects.

[1] http://www.birds.cornell.edu/citscitoolkit/evaluation/instruments

REFERENCES

1. Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis* 20, 3 (2012), 351–368. DOI: http://dx.doi.org/10.1093/pan/mpr057
2. Jeremy P. Birnholtz and Matthew J. Bietz. 2003. Data at Work: Supporting Sharing in Science and Engineering. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work (GROUP '03)*. ACM, New York, NY, USA, 339–348. DOI: http://dx.doi.org/10.1145/958160.958215
3. David Blumenthal, Eric G Campbell, Manjusha Gokhale, Recai Yucel, Brian Clarridge, Stephen Hilgartner, and Neil a Holtzman. 2006. Data withholding in genetics and the other life sciences: prevalences and predictors. *Academic Medicine* 81, 2 (2006), 137–145. DOI: http://dx.doi.org/10.1097/00001888-200602000-00008
4. Sara Bly. 1998. Special Section on Collaboratories. *interactions* 5, 3 (May 1998), 31. DOI: http://dx.doi.org/10.1145/275269.275273
5. John Bohannon. 2016. Psychologists grow increasingly dependent on online research subjects. (2016). Retrieved June 7, 2016 from http://www.sciencemag.org/news/2016/06/psychologists-grow-increasingly-dependent-online-research-subjects.
6. Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 11 (2009), 977–984. DOI: http://dx.doi.org/10.1525/bio.2009.59.11.9
7. Anne Bowser, Derek Hansen, Jennifer Preece, Yurong He, Carol Boston, and Jen Hammock. 2014. Gamifying Citizen Science: A Study of Two User Groups. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work (CSCW Companion '14)*. ACM, New York, NY, USA, 137–140. DOI: http://dx.doi.org/10.1145/2556420.2556502
8. Anne Bowser-Livermore and Andrea Wiggins. 2015. Privacy in Participatory Research: Advancing Policy to support Human Computation. *Human Computation* 2, 1 (2015), 19–44. DOI:http://dx.doi.org/10.15346/hc.v2i1.3
9. Alex Bowyer, Chris Lintott, Greg Hines, Campbell Allan, and Ed Paget. 2015. Panoptes, a Project Building Tool for Citizen Science. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP '15)*. AAAI, San Diego, CA, USA, 1–2.
10. H.K. Burgess, L.B. DeBey, H.E. Froehlich, N. Schmidt, E.J. Theobald, A.K. Ettinger, J. HilleRisLambers, J. Tewksbury, and J.K. Parrish. 2016. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation* (2016), 1–8. DOI: http://dx.doi.org/10.1016/j.biocon.2016.05.014
11. Carolin Cardamone, Kevin Schawinski, Marc Sarzi, Steven P. Bamford, Nicola Bennert, C. M. Urry, Chris Lintott, William C. Keel, John Parejko, Robert C. Nichol, Daniel Thomas, Dan Andreescu, Phil Murray, M. Jordan Raddick, Anze Slosar, Alex Szalay, and Jan Vandenberg. 2009. Galaxy Zoo Green Peas: Discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society* 399, 3 (2009), 1191–1205. DOI: http://dx.doi.org/10.1111/j.1365-2966.2009.15383.x
12. Casey Trees. 2016. Citizen Science. (2016). Retrieved January 15, 2016 from http://caseytrees.org/get-involved/citizen-science/.
13. Karin Knorr Cetina. 1999. *Epistemic Communities: How the Sciences Make Knowledge*. Harvard Education Press, Cambridge, MA.
14. Sayan Chakraborti, Naveen Yadav, Carolin Cardamone, and Alak Ray. 2012. Radio Detection of Green Peas: Implications for Magnetic Fields in Young Galaxies. *The Astrophysical Journal* 746, 1 (2012), L6. DOI: http://dx.doi.org/10.1088/2041-8205/746/1/L6
15. Jeffrey P Cohn. 2008. Citizen Science : Can Volunteers Do Real Research? *BioScience* 58, 3 (2008), 192–197. DOI:http://dx.doi.org/10.1641/B580303
16. Harry Collins. 1992. *Changing Order: Replication and Induction in Scientific Practice*. University of Chicago Press, Chicago, IL.
17. Harry Collins and Robert Evans. 2007. *Rethinking Expertise*. University of Chicago Press.
18. Colorado State University. 2016. Citsci.Org. (2016). Retrieved January 15, 2016 from http://www.citsci.org.
19. Caren Cooper. 2016. Zen in the Art of Citizen Science: Apps for Collective Discovery and the 1 Percent Rule of the Web. (2016). Retrieved January 15, 2016 from http://blogs.scientificamerican.com/guest-blog/zen-in-the-art-of-citizen-science-apps-for-collective-discovery-and-the-1-percent-rule-of-the-web/.
20. Juliet M. Corbin and Anselm Strauss. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd ed.). Sage, Thousand Oaks, CA.
21. M. L. Cornwell and L. M. Campbell. 2012. Co-producing conservation and knowledge: Citizen-based sea turtle monitoring in North Carolina, USA. *Social Studies of Science* 42, 1 (2012), 101–120. DOI: http://dx.doi.org/10.1177/0306312711430440
22. Joe Cox, Eun Young Oh, Brooke Simmons, Chris Lintott, Karen Masters, Anita Greenhill, Gary Graham, and Kate Holmes. 2015. Defining and Measuring Success in Online Citizen Science: A Case Study of Zooniverse Projects. *IEEE Computing in Science & Engineering* 17, 4 (2015), 28–41. DOI:http://dx.doi.org/10.1109/MCSE.2015.65

23. M. H. Cragin, C. L. Palmer, J. R. Carlson, and M. Witt. 2010. Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, 1926 (2010), 4023–4038. DOI: http://dx.doi.org/10.1098/rsta.2010.0165

24. Schlagwein Daniel and Daneshgar Farhad. 2014. User requirements of a crowdsourcing platform for researchers : findings from a series of focus groups. In *Proceedings of the 2014 Pacific Asia Conference on Information Systems*. 1–11.

25. Janet Dickinson and Rick Bonney. 2012. *Citizen Science : Public Participation in Environmental Research*. Cornstock Pub. Associates, Ithaca, NY.

26. Michael Dummett. 1991. *The Logical Basis of Metaphysics*. Duckworth, London, UK.

27. Kevin Dunbar. 2000. How Scientists Think in the Real World. *Journal of Applied Developmental Psychology* 21, 1 (2000), 49–58. DOI: http://dx.doi.org/10.1016/S0193-3973(99)00050-7

28. Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L. Cox. 2014. Designing for Dabblers and Deterring Drop-outs in Citizen Science. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2985–2994. DOI: http://dx.doi.org/10.1145/2556288.2557262

29. Elizabeth J Farnsworth, Miyoko Chu, W John Kress, Amanda K Neill, Jason H Best, John Pickering, Robert D Stevenson, Gregory W Courtney, John K V Dyk, and Aaron M Ellison. 2013. Next-Generation Field Guides. *BioScience* 63, 11 (2013), 891–899. DOI: http://dx.doi.org/10.1525/bio.2013.63.11.8

30. Karen Fort, Gilles Adda, and K. Bretonnel Cohen. 2011. Amazon Mechanical Turk: Gold Mine or Coal Mine? *Computational Linguistics* 37, 2 (June 2011), 413–420. DOI:http://dx.doi.org/10.1162/COLI_a_00057

31. Amy Freitag and Max Pfeffer. 2013. Process, Not Product: Investigating Recommendations for Improving Citizen Science Success. *PLoS One* 8, 5 (2013), e64079. DOI:http://dx.doi.org/10.1371/journal.pone.0064079

32. Bruno S. Frey and Reto Jegen. 2001. Motivation Crowding Theory. *Journal of Economic Surveys* 15, 5 (2001), 589–611. DOI: http://dx.doi.org/10.1111/1467-6419.00150

33. Jay R Galbraith. 1974. Organization Design: An Information Processing View. *Interfaces* 4, 3 (1974), 28–36. DOI:http://dx.doi.org/10.1287/inte.4.3.28

34. Thomas F. Giergyn. 1999. *Cultural Boundaries of Science: Credibility on the Line*. University of Chicago Press, Chicago, IL.

35. Thomas F Gieryn. 1983. Boundary-Work and the Demarcation of Science From Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review* 48, 6 (1983), 781–795. DOI:http://dx.doi.org/Doi10.2307/2095325

36. Elizabeth Goodman and Janet Vertesi. 2012. Design for X?: Distribution Choices and Ethical Design. In *Extended Abstracts of the 30th Annual ACM Conference on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 81–90. DOI: http://dx.doi.org/10.1145/2212776.2212786

37. David Grier. 2007. *When Computers Were Humans*. Princeton University Press, Princeton, NJ.

38. Katie Hafner and Matthew Lyon. 1996. *Where Wizards Stay up Late: the Origins of the Internet*. Simon & Schuster, New York.

39. Muki Haklay. 2013. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer, Berlin, Chapter Citizen Science and Volunteer Geographic Information: Overview and Typology of Participation, 105–122. DOI: http://dx.doi.org/10.1007/978-94-007-4587-2_7

40. Harvard School of Engineering and Applied Sciences. 2016. Privacy Tools for Sharing Research Data. (2016). Retrieved January 15, 2016 from http://privacytools.seas.harvard.edu.

41. Imperial College London. 2016. EpiCollect.net Website. (2016). Retrieved January 15, 2016 from http://www.epicollect.net.

42. Yuri I. Izotov, Natalia G. Guseva, and Trinh X. Thuan. 2011. Green Pea Galaxies and Cohorts: Luminous Compact Emission-Line Galaxies in the Sloan Digital Sky Survey. *The Astrophysical Journal* 728, 2 (2011), 161. DOI:http://dx.doi.org/10.1088/0004-637X/728/2/161

43. Corey Jackson, Gabriel Mugar, Kevin Crowston, and Carsten Osterlund. 2016. Encouraging Work in Citizen Science: Experiments in Goal Setting and Anchoring. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, 297–300. DOI: http://dx.doi.org/10.1145/2818052.2869129

44. A. E. Jaskot and M.S. Oey. 2013. The Origin and Optical Depth of Ionizing Radiation in the "Green Pea" Galaxies. *The Astrophysical Journal* 766 (2013), 91. DOI: http://dx.doi.org/10.1088/0004-637X/766/2/91

45. Marina Jirotka, Charlotte P. Lee, and Gary M. Olson. 2013. Supporting scientific collaboration: Methods, tools and concepts. *Computer Supported Cooperative Work: CSCW: An International Journal* 22, 4-6 (2013), 667–715. DOI:http://dx.doi.org/10.1007/s10606-012-9184-0

46. Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. 2013. Sensr: Evaluating a Flexible Framework for Authoring Mobile Data-collection Tools for Citizen Science. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1453–1462. DOI: http://dx.doi.org/10.1145/2441776.2441940

47. Thomas Kuhn. 2012. *The Structure of Scientific Revolutions* (4th ed.). University of Chicago Press, Chicago, IL.

48. Deepak Kulkarni and Herbert A. Simon. 1988. The processes of scientific discovery: The strategy of experimentation. *Cognitive Science* 12, 2 (1988), 139–175. DOI: http://dx.doi.org/10.1016/0364-0213(88)90020-1

49. Bruno Latour and Steve Woolgar. 1986. *Laboratory life: the Social Construction of Scientific Facts* (2nd ed.). Princeton University Press, Beverly Hills, CA.

50. Edith Law, Connor Dalton, Nick Merrill, Alex Young, and Krzysztof Z. Gajos. 2013. Curio: a platform for supporting mixed-expertise crowdsourcing. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP '13)*. AAAI, 1–2.

51. Chris J. Lintott, Kevin Schawinski, Anze Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. 2008. Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389, 3 (2008), 1179–1189. DOI: http://dx.doi.org/10.1111/j.1365-2966.2008.13689.x

52. Christopher S. Lowry and Michael N. Fienen. 2013. CrowdHydrology: Crowdsourcing hydrologic data and engaging citizen scientists. *GroundWater* 51, 1 (2013), 151–156. DOI: http://dx.doi.org/10.1111/j.1745-6584.2012.00956.x

53. Marder, Jenny and Fritz, Mike. (????).

54. Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (2012), 1–23. DOI: http://dx.doi.org/10.3758/s13428-011-0124-6

55. R. Merton, E. Barber, and J. Shulman. 2006. *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton University Press.

56. New York University. 2016. Project Noah. (2016). Retrieved January 15, 2016 from http://www.projectnoah.org.

57. Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z. Gajos. 2011. Platemate: Crowdsourcing Nutritional Analysis from Food Photographs. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 1–12. DOI: http://dx.doi.org/10.1145/2047196.2047198

58. Gary Olson, Ann Zimmerman, Nathan Bos, and William Wulf. 2008. *Scientific Collaboration on the Internet*. MIT Press, Cambridge, MA.

59. Gary M. Olson, Daniel E. Atkins, Robert Clauer, Thomas A. Finholt, Farnam Jahanian, Timothy L. Killeen, Atul Prakash, and Terry Weymouth. 1998. The Upper Atmospheric Research Collaboratory (UARC). *interactions* 5, 3 (May 1998), 48–55. DOI: http://dx.doi.org/10.1145/275269.275276

60. Johan Oomen and Lora Aroyo. 2011. Crowdsourcing in the Cultural Heritage Domain : Opportunities and Challenges. *C&T '11 Proceedings of the 5th International Conference on Communities and Technologies* July (2011), 138–149. DOI: http://dx.doi.org/10.1145/2103354.2103373

61. Michael Polanyi. 1967. *The Tacit Dimension*. Routledge and Kegan Paul, London, UK.

62. J. Preece and B. Schneiderman. 2009. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction* 1, 1 (2009), 13–32. DOI: http://dx.doi.org/10.5121/ijfcst.2014.4403

63. Nancy Proctor. 2013. Crowdsourcing–An Introduction: From Public Goods to Public Good. *Curator: The Museum Journal* 56, 1 (January 2013), 105—106. DOI: http://dx.doi.org/DOI:10.1111/cura.12010

64. Danial Qaurooni, Ali Ghazinejad, Inna Kouper, and Hamid Ekbia. 2016. Citizens for Science and Science for Citizens: The View from Participatory Design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1822–1826. DOI: http://dx.doi.org/10.1145/2858036.2858575

65. M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 2010. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review* 9, 1 (2010), 010103. DOI: http://dx.doi.org/10.3847/AER2009036

66. Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work (CSCW '15)*. ACM, New York, NY, USA, 1364–1378. DOI: http://dx.doi.org/10.1145/2675133.2675246

67. Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 75–85. DOI: http://dx.doi.org/10.1145/2642918.2647409

68. Hauke Riesch and Clive Potter. 2014. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science* 23, 1 (2014), 107–120. DOI: http://dx.doi.org/10.1177/0963662513497324

69. Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic Changes in Motivation in Collaborative Citizen-science Projects. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 217–226. DOI: http://dx.doi.org/10.1145/2145204.2145238

70. Sharman Apt Russell. 2014. *Diary of a Citizen Scientist*. Oregon State University Press, Corvallis, OR.

71. SciFabric. 2016a. CrowdCrafting.org Website. (2016). Retrieved January 15, 2016 from http://crowdcrafting.org.

72. SciFabric. 2016b. Pybossa. (2016). Retrieved January 15, 2016 from http://www.pybossa.com.

73. SciStarter. 2016. Project Finder. (2016). Retrieved January 15, 2016 from http://scistarter.com/finder.

74. S. Shapin. 1995. Here and Everywhere: Sociology of Scientific Knowledge. *Annual Review of Sociology* 21, 1 (1995), 289–321. DOI: http://dx.doi.org/10.1146/annurev.soc.21.1.289

75. Jennifer Shirk. 2014. *Push the Edge of Science Forward: Expanding Considerations of Expertise Through Scientists' Citizen Science Work in Conservation*. Ph.D. Dissertation. Cornell University, Ithaca, NY.

76. Jennifer L. Shirk, Heidi L. Ballard, Candie C. Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, Matthew Minarchek, Bruce V. Lewenstein, Marianne E. Krasny, and Rick Bonney. 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and Society* 17, 2 (2012). DOI: http://dx.doi.org/org/10.5751/ES-04705-170229

77. Jonathan Silvertown. 2009. A new dawn for citizen science. *Trends in ecology & evolution* 24, 9 (2009), 467–71. DOI:http://dx.doi.org/10.1016/j.tree.2009.03.017

78. Robert Simpson, Kevin R. Page, and David De Roure. 2014. Zooniverse: Observing the World's Largest Citizen Science Platform. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 1049–1054. DOI:http://dx.doi.org/10.1145/2567948.2579215

79. Ryo Suzuki, Niloufar Salehi, Michelle S. Lam, Juan C. Marroquin, and Michael S. Bernstein. 2016. Atelier: Repurposing Expert Crowdsourcing Tasks As Micro-internships. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2645–2656. DOI: http://dx.doi.org/10.1145/2858036.2858121

80. Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6, 6 (2011). DOI: http://dx.doi.org/10.1371/journal.pone.0021101

81. Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Rösch, Robert Simpson, and Nigel Shadbolt. 2015. Designing for Citizen Data Analysis: A Cross-Sectional Case Study of a Multi-Domain Citizen Science Platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4069–4078. DOI: http://dx.doi.org/10.1145/2702123.2702420

82. Sharon Traweek. 1988. *Beamtimes and Lifetimes: the World of High Energy Physicists*. Harvard University Press, Cambridge, MA.

83. Michael L Tushman and David A Nadler. 1978. Information Processing as an Integrating Concept in Organizational Design. *Source: The Academy of Management Review* 3, 3 (1978), 613–624. DOI: http://dx.doi.org/10.5465/AMR.1978.4305791

84. Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185, 4157 (1974), 1124–1131. DOI: http://dx.doi.org/10.1126/science.185.4157.1124

85. University College of London. 2016. Transcribe Bentham: A Participatory Initiative. (2016). Retrieved January 15, 2016 from http://blogs.ucl.ac.uk/transcribe-bentham/.

86. Andrew J Westphal, Anna L Butterworth, Christopher J Snead, Nahide Craig, David Anderson, Steven M Jones, Donald E Brownlee, Richard Farnsworth, and Michael E Zolensky. 2005. Stardust@home: A Massively Distributed Public Search for Interstellar Dust in the Stardust Interstellar Dust Collector. In *Lunar and Planetary Science XXXVI*. http://ntrs.nasa.gov/search.jsp?R=20050180792

87. Andrea Wiggins. 2012. *Crowdsourcing Scientific Work: A Comparative Study of Technologies, Processes, and Outcomes in Citizen Science*. Ph.D. Dissertation. Syracuse University, Syracuse, NY.

88. Andrea Wiggins and Kevin Crowston. 2011. From Conservation to Crowdsourcing: A Typology of Citizen Science. In *Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS '11)*. IEEE Computer Society, Washington, DC, USA, 1–10. DOI: http://dx.doi.org/10.1109/HICSS.2011.207

89. Andrea Wiggins and Kevin Crowston. 2012. Goals and Tasks: Two Typologies of Citizen Science Projects. In *Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS '12)*. IEEE Computer Society, Washington, DC, USA, 3426–3435.

90. A. Wiggins, G. Newman, R.D. Stevenson, and K. Crowston. 2011. Mechanisms for Data Quality and Validation in Citizen Science. In *IEEE seventh International Conference on e-Science Workshops (eScienceW '11)*. 14–19.

91. Alex C Williams, John F Wallin, Haoyu Yu, Marco Perale, Hyrum D Carroll, Anne-Francoise Lamblin, Lucy Fortson, Dirk Obbink, Chris J Lintott, and James H Brusuelas. 2014. A computational pipeline for crowdsourced transcriptions of Ancient Greek papyrus fragments. In *Proceedings of 2014 IEEE International Conference on Big Data*. 100–105. DOI: http://dx.doi.org/10.1109/BigData.2014.7004460

92. Chris Wood, Brian Sullivan, Marshall Iliff, Daniel Fink, and Steve Kelling. 2011. eBird: Engaging birders in science and conservation. *PLoS Biology* 9, 12 (2011). DOI:http://dx.doi.org/10.1371/journal.pbio.1001220

93. A Zimmerman. 2008. New knowledge from old data - The role of standards in the sharing and reuse of ecological data. *Science Technology & Human Values* 33, 5 (2008), 631–652. DOI: http://dx.doi.org/10.1177/0162243907306704

94. Ann Zimmerman, Bos Nathan, Judy Olson, and Gary Olson. 2009. *e-Research: Transformation in Scholarly Practice*. Routledge, Chapter The Promise of Data in e-Research: Many Challenges, Multiple Solutions, Diverse Outcomes, 222–239.

95. Zooniverse. 2016a. AnnoTate. (2016). Retrieved January 15, 2016 from https://anno.tate.org.uk/.

96. Zooniverse. 2016b. Old Weather Project. (2016). Retrieved January 15, 2016 from http://www.oldweather.org/.

97. Zooniverse. 2016c. Operation War Diary. (2016). Retrieved January 15, 2016 from http://www.operationwardiary.org/.