Digestif: Promoting Science Communication in Online Experiments

EUNICE JUN, Computer Science & Engineering, DUB Group, University of Washington
BLUE JO, Computer Science & Engineering, DUB Group, University of Washington
NIGINI OLIVEIRA, Computer Science & Engineering, DUB Group, University of Washington
KATHARINA REINECKE, Computer Science & Engineering, DUB Group, University of Washington

Online experiments allow researchers to collect data from large, demographically diverse global populations. Unlike in-lab studies, however, online experiments often fail to inform participants about the research to which they contribute. This paper is the first to investigate barriers that prevent researchers from providing such science communication in online experiments. We found that the main obstacles preventing researchers from including such information are assumptions about participant disinterest, limited time, concerns about losing anonymity, and concerns about experimental bias. Researchers also noted the dearth of tools to help them close the information loop with their study participants. Based on these findings, we formulated design requirements and implemented Digestif, a new web-based tool that supports researchers in providing their participants with science communication pages. Our evaluation shows that Digestif's scaffolding, examples, and nudges to focus on participants make researchers more aware of their participants' curiosity about research and more likely to disclose pertinent research information.

CCS Concepts: • **Collaborative and social computing systems and tools**;

ACM Reference format:
Eunice Jun, Blue Jo, Nigini Oliveira, and Katharina Reinecke. 2018. Digestif: Promoting Science Communication in Online Experiments. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 84 (November 2018), 26 pages.
https://doi.org/10.1145/3274353

1 INTRODUCTION

Over the past two decades, online experiments have become increasingly popular for research in various fields [6, 12, 15, 21, 28, 44]. Compared to laboratory experiments, online studies offer faster, more effortless, and less costly participant recruitment and larger, more diverse samples [6, 24, 29, 39, 45]. As a consequence, they enhance sample representativeness and the generalizability of findings. However, guidelines for the ethical treatment of participants in the lab have rarely been applied to the online setting. Unlike in-lab studies where participants contribute data and learn how their data will be used to achieve a research aim, participants in online experiments often receive little information about the research they just participated in [49, 50]. Many IRBs even exempt online studies conducted on Amazon's Mechanical Turk (MTurk) from informed consent because they are considered low risk [43].

We believe that describing study aims and soliciting participant feedback are ethical imperatives when conducting online experiments. In fact, the American Psychological Association (APA)

2573-0142/2018/11-ART84 $15.00
https://doi.org/10.1145/3274353

specifies that researchers provide all study participants, in and out of the lab, with both pre- and post-experiment feedback [2].

One place where participants are more informed of the impact of their data is in volunteer-based online experiments on platforms such as LabintheWild [44], TestMyBrain [21], and GamesWithWords [57]. Most volunteer-based online experiments compensate participants with *conclusion pages* that tell them about their performance and the research aims after completing an experiment. A subset of the content found on conclusion pages is in line with the APA's guidelines on post-experiment feedback and can be used to promote science communication. Drawing from literature that synthesizes definitions across disciplines [14], we define science communication to be information that gives participants context about the research they contribute to, such as information about the experiment's design, research goals, hypotheses, and previous work in the area, as well as opportunities to engage with researchers to ask questions or resolve concerns.

There is convincing evidence that participants appreciate such additional information. For example, although the primary motivation for participants on MTurk is monetary gain [29, 46], researchers found that 40% of participants taking experiments on MTurk later sought out further information on the experiment's research topic and design without being paid [32].

In this paper, we begin to re-imagine large scale online experimentation to support researchers in becoming more aware of their participants' interests and to promote mutual understanding and communication, aspects necessary for the sustained growth of crowdsourcing [34].

Our aim is to promote the inclusion of science communication pages at the end of online experiments—a goal that necessitates understanding and mitigating the barriers researchers face when providing such information to their study participants. We began with two formative studies: (1) an analysis of what kinds of information researchers already provide to volunteer participants, defining the design space of current science communication pages in online experiments, and (2) interviews with two groups of online experiment researchers, those who do and do not include post-experiment research information. The second formative study investigated why researchers provide their participants with different aspects of the design space of science communication pages and any challenges they face in doing so. Synthesizing findings from the two formative studies, we identified design guidelines. To validate our design guidelines, we built and evaluated Digestif, a web-based system that supports researchers in providing post-experiment science communication. Table 1 shows an overview of the multiple phases included in this paper.

This paper makes three main contributions:

(1) Our work focuses on researchers' challenges for closing the information loop with research participants and extends prior work that focuses on participant needs [30, 42, 49, 50]. We found that the main obstacles preventing researchers from including science communication information are a lack of awareness of their participants' interests, limited time, issues pertaining to their own privacy, concerns about experiment bias, and a dearth of tools to support creating these pages.
(2) We offer design guidelines that promote ethically grounded online study practices – such as surfacing participant interests and their feedback and providing opportunities for communicating research context to participants.
(3) We explore and implement these design guidelines through Digestif, a publicly available, web-based system that supports researchers in creating science communication pages for their online experiments. Our evaluation validates the design guidelines and system. We found that Digestif mitigates the barriers we identified in the formative studies, making researchers more aware of their participants' curiosity about research and increasing their interest in providing additional information in the future.

Table 1. We conducted two formative studies to derive design requirements for systems to support creating science communication pages in online experiments. The design requirements were then validated using Digestif, a novel system for supporting researchers in providing such pages. This table outlines the number of participants involved, purpose, and main findings from each research step presented in this paper.

	Formative Studies		Exploratory System	
	Landscape Analysis	**Interviews**	**Digestif**	**Evaluation**
	26 conclusion pages, 11 online experiments	13 researchers		14 researchers with Digestif, 9 without Digestif
Purpose	To analyze the breadth of information researchers provide their participants	To identify the barriers and challenges researchers face to providing research information	To explore and implement the design requirements	To evaluate Digestif and extract further design insights
Findings	• Science communication is possible in various kinds of online experiments. • 10 key ways for researchers to communicate with participants (Table 2) — later used to scaffold Digestif • Examples for library in Digestif • Questions for semi-structured interviews (second formative study) • Design requirements	• Key barriers and challenges • Researchers' current perspectives on participants • Design requirements	• Exploratory implementation of the design requirements in Digestif (Section 5) • Tensions and insights that informed ultimately led to key design principles (Discussion)	• Example-driven scaffolding is key • Nudges that portray participants increase awareness of participants • Key design principles • (see Discussion)

2 BACKGROUND AND RELATED WORK

Our research builds on existing work on the ethical importance of participant awareness in experiments, practices found in volunteer-based online experiment platforms, and participant needs for learning opportunities.

Ethics in Laboratory and Online Studies. The Belmont Report outlines three basic ethical principles for human subjects research: respect for persons (honoring the autonomy of participants and protecting those who may not be able to act autonomously), beneficence (do no harm to participants but also ensure that the possible benefits outweigh the risks), and justice (equal recruitment of participants and equal distribution of benefits from research to society) [18]. Based on the Belmont report, the APA developed a code of conduct to protect participants and researchers [2]. It outlines practices such as pre-observational informed consent procedures, which include information about the research aim, and post-observational debriefings, which provide background information about experiment surveys and tasks [54]. Institutional Review Boards (IRBs) enforce many of these procedures.

Researchers have raised concerns that these in-lab procedures to protect and inform participants may not have transferred to online studies [9, 11, 35, 56]. While many ethics guidelines for online studies exist (see, e.g., [3, 10, 19, 51]), adherence is uncertain [26]. This is mostly because "ensuring informed consent, explaining instructions, and conducting effective debriefings online may be more difficult than in traditional laboratory settings" [35]. Indeed, Jun et al. found that only 10% of voluntary participants closely read the informed consent page in online studies [33]. In a compensated setting where participants are incentivized to complete tasks as quickly as possible, the percentage will likely be lower. Debriefing participants in an anonymous online setting is equally difficult, if not impossible due to scale. It is nontrivial to assess people's reactions to the study or the research experience [26, 35]. Many IRBs have declared some studies with anonymous online participants as exempt, notably often those conducted on the online labor market MTurk [43]. Many online studies conducted on MTurk (or similar platforms, such as Crowd Flower) therefore do not collect informed consent or provide any form of debriefing.

Although guidelines for conducting online experiments suggest researchers provide more contextual information in online studies, e.g., by enabling online participants to learn about the results of their studies [26], research studies conducted on MTurk rarely do so. As a result, there is an "information asymmetry" between requesters (i.e., those who post a task or study) and workers on MTurk; workers often do not know what research goals their work supports [49]. As a response to such concerns, researchers have developed Tukopticon [30], a website workers can use to rate and review requesters and share information with each other, and Daemo [20], a crowd-designed online labor market that involved crowd workers in its design, development, and launch. Additional guidelines, including suggestions (1) to render researchers and participants more visible and (2) to give participants more context for their work, for both system designers and researchers have also emerged [4]. All these approaches give voice to and prioritize participants' needs, but none inquire into researchers' challenges and perspectives in online experiments.

Our work is the first to bring researchers' practical challenges and views into the conversation about online research ethics.

Participant Learning Opportunities in Online Studies and in Crowd Work. Volunteer-based online experiment platforms, such as MySocialBrain [40], LabintheWild [36], and GamesWithWords [57] provide some information that would be exchanged at the end of in-lab studies on the last page ("conclusion page"). In contrast to debriefing pages, which reveal the true, previously concealed nature of a study after participants' behavior is observed [54], conclusion pages often provide additional learning opportunities and the chance for participants to share their thoughts with the researchers. Rather than compensate participation with money, volunteer-based online experiments compensate participants with information about the research, their performance, or a personal characteristic (e.g., thinking style or writing style). Recent work has found that volunteer participants are eager to learn more about themselves, the research they contribute to, and the study design [42].

Additionally, research has shown that both volunteer and financially compensated participants learn about the research being conducted from online studies and any research information provided at the end. For example, online studies can provide casual observational learning throughout participation [13].MTurk participants have voluntarily sought out, spent unpaid time on and learned from pages with research information, and positively responded to the presented material and opportunities for providing feedback [32].

Informed consent forms provide a glimpse of the study and broader research goals, but they are not enough. Participants in laboratory studies receive more extensive information through informal conversations with researchers that are often guided by participants' curiosity. For example, before a participant leaves a lab, best practice guidelines suggest that researchers should ask whether participants have any more questions, in line with the APA debriefing guideline to provide "a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research" [2]. Since online participants give data that would otherwise have to be collected in the lab, they should also receive this information.

Closing the communication loop between participants and researchers by providing research information will benefit both researchers and participants through more ethical online experiments and additionally benefit participants with learning opportunities.

Table 2. Study 1: Types of content blocks we identified in conclusion pages. * indicates information that supports the two facets of science communication: research context and participant opportunities to engage with researchers.

Content Block	**Explanation**
Acknowledgement	Thank participants for their time and data contribution
Community Building*	Show participants how to connect with researchers, e.g.,joining a mailing list or liking a study platform on social media
Experimental Design*	Tell participants why the online experiment was designed the way it was; Explain how any personalized results were generated and limitations of the study
Other Studies	Contain links to other studies on a platform, leverage current traffic to attract more participation to other studies
Participant Feedback*	Allow participants to provide suggestions, ideas, and concerns about the tasks. Based on the feedback, researchers can refine their studies iteratively
Personalized Results	Provide participants some insight into their performance/data. A common form of personalized results is social comparison
Previous Research*	Share about researchers' fields and previous research findings
Research Goals*	Explain how the results of the study will contribute to advancing various disciplines or improving applications (e.g., possible products, etc.)
Research Motivations*	Tell participants why data is being collected and/or the hypotheses that the study addresses
Share	Invite participants to share their participation on social media, which can generate traffic for the study

3 FORMATIVE STUDY 1: WHAT COULD BE INCLUDED IN SCIENCE COMMUNICATION PAGES?

Volunteer-based online experiments include conclusion pages that provide research information, but similar pages do not exist in financially compensated online experiments. To systematically understand what could be included in pages dedicated to science communication, we collected and analyzed conclusion pages for the research information they include.

When we use the term "online experiments," we refer to controlled studies, surveys, and other tasks used for understanding human behavior. We do not refer to annotation tasks that are used solely for collecting large datasets (e.g., image labeling).

3.1 Methods

We used a purposive sampling method, "landscape sampling," following the procedure described in [7] and [55]. While landscape sampling did not allow us to generalize or draw conclusions about the frequency of certain content types that researchers provide after their studies, it suited our goal for widely exploring the possible kinds of information they include in conclusion pages by identifying as many different content types and designs as possible.

We began our search for studies that provide participants with post-experiment research information on the Social Psychology Network [41] and SciStarter [47], two publicly available lists of online studies and citizen science projects. Once we identified a few online studies, we contacted the researchers conducting them and asked for recommendations of other, similar projects, following a snowball recruiting technique. We included online studies if they were not already in our dataset and included research content that was new or presented in a novel way.

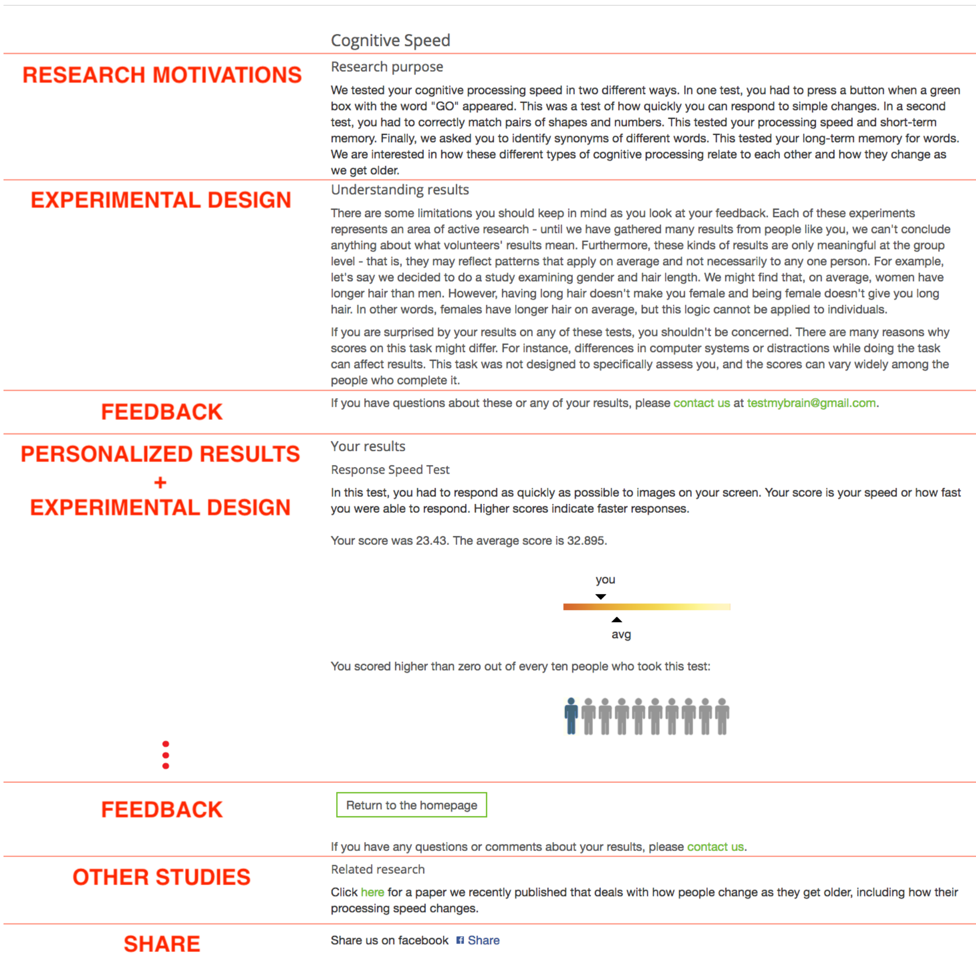

Fig. 1. Sample conclusion page with its content labeled. See Table 2 for descriptions of the content types. The personalized results and experimental design information has been shortened in this figure.

3.2 Dataset

We found a small number of online study platforms or websites that provide conclusion pages. This practice is largely restricted to online studies that rely on volunteer participants. Our final sample includes conclusion pages from 26 studies hosted on 11 different online study platforms or websites. Table 3 gives an overview of the disciplines represented in our conclusion pages. Figure 1 is a conclusion page we included in our landscape sample. We make our dataset publicly available at http://digestif.labinthewild.org/digestif/explore.

3.3 Analysis

After an initial pass through the landscape sample, the first author developed high level codes based on the types of content found on conclusion pages, the layout of the content, when the pages were provided, the fields of study the pages came from, the kind of study (i.e., survey or experiment), and any noticeable tones or appeals to motivations. Two authors then used an inductive coding procedure.

Table 3. Study 1 dataset: We sampled conclusion pages from experiments conducted in ten different disciplines.

Discipline	**Count (%)**
Psychology	8 (31%)
Cognitive science	4 (15%)
HCI	3 (12%)
Social science	3 (12%)
Neuroscience	2 (8%)
Computer science	2 (8%)
Animal science	1 (4%)
Behavioral economics	1 (4%)
Linguistics	1 (4%)
Nutrition	1 (4%)
Total	**26**

Both coded one page, compared codes, and discussed to reach consensus before coding all conclusion pages in the dataset and discussing any emergent codes or disagreements along the way. To extract insights from the codes, the authors conducted a thematic analysis using affinity diagrams [8].

3.4 Key Findings

Our analysis revealed ten main content blocks, or kinds of information, found on conclusion pages (Table 2). Of the ten, six specific forms of information served science communication purposes. Information about the *experimental design* of an online study, *previous research* on the topic, long-term *research goals*, and *research motivations* for a particular online study could all increase participants' understanding of a science domain and the scientific process; asking participants for *feedback* and to join a *community* to engage with the researchers longer term could lead to greater communication and understanding between participants and researchers.

The conclusion pages varied along four main axes: content block order, linguistic style, types of media included, and timing of delivery.

3.4.1 Number and Order of Content Blocks. Twenty (of 26) scientific outreach pages in our sample had between two to eight unique content blocks. The simplest conclusion page in our sample included only a personalized result, but this page was an anomaly on the platform and in our dataset. 15 conclusion pages in our sample featured at least six blocks of content. These pages required significant scrolling but conveyed a wider range of information, which likely met more participants' interests. In our dataset, the same information often appeared twice on a single conclusion page in summary and detailed form.

3.4.2 Linguistic Style of Conclusion Pages. Most post-experiment informative pages conveyed an overall positive, excited, and informal but authoritative tone. Participants were usually addressed in second-person and researchers in third-person ("we"), which created a sense of conversation and familiarity between participants and researchers. The popularity of a conversational tone suggests that conclusion pages may be places where dialogue is possible, as compared to during experimental tasks. Other pages were more formal, straightforward, and impersonal in tone.

3.4.3 Types of Media Included. The vast majority of the conclusion pages in our sample were self-contained and only relied on external links to take participants to future studies. However, four pages in our sample used hyperlinks to online articles and resources to illustrate concepts in more familiar ways to participants. Two of these included YouTube videos introducing key concepts related to the online studies. The creative usage of media online suggests that conclusion pages provide opportunities that go beyond the conversations participants and researchers could have in the lab. Conclusion pages are unique opportunities for participants to engage in science and researchers to communicate their work.

3.4.4 Timing of Conclusion Pages. Most studies showed participants all blocks of the conclusion page immediately after completing the study. However, studies occasionally delayed some information, offering to email the overall results of the study to participants or providing a link to a social media site where results would be posted in the future. The delays appeared to be more pragmatic, such as when research results may not be available yet. Still, they could encourage participants to become more involved by following a lab on social media, for instance.

4 FORMATIVE STUDY 2: WHY DO SOME RESEARCHERS PROVIDE RESEARCH INFORMATION WHILE OTHERS DO NOT?

From our first formative study we learned that researchers can communicate a broad range of information in an authoritative yet conversational tone, providing information immediately after a study or at some future point and using simple or multimedia presentations. Exploring the design space of conclusion pages raised many additional questions around *why* some content blocks were provided on conclusion pages and *how* researchers came to make specific design decisions.

To answer these questions, we interviewed researchers who do and do not provide participants with research information. Our aim was to understand their motivations for sharing or not sharing information, processes of creating conclusion pages, and challenges to providing participants with science communication pages.

Four research questions guided our interviews and analysis:

(1) Why do researchers include certain kinds of science information at the end of studies?
(2) What is the workflow researchers use to create conclusion pages?
(3) What challenges do researchers face in creating conclusion pages?
(4) What prevents researchers who have never provided additional information about their research from including science communication pages in their online experiments?

4.1 Method

We conducted semi-structured interviews with 13 researchers from diverse fields, including medicine, psychology, and computer science. Seven of these researchers had previously used conclusion pages. The others frequently conducted online experiments but had not previously included any additional information on their research. We recruited people by contacting researchers from the platforms we found in our landscape sample and used snowball recruitment until our findings converged. We conducted the interviews in-person (N=2), over Skype (N=10), and via email (N=1). Skype and in-person interviews were recorded and transcribed verbatim. The interview structure was informed by our landscape sample and research questions. The interviews lasted approximately 30 minutes each. Our IRB exempted this study from full review.

We interviewed researchers who use conclusion pages to probe into their motivations for including certain kinds of science information, process for developing the pages, their knowledge of and interactions with participants, and their imagination of an ideal workflow to provide participants with additional information about their participation in online experiments. Researchers were asked to provide links to their studies beforehand to ground the conversation. We additionally shared our screen showing their conclusion page when applicable.

With researchers who had not provided their participants with any additional information before, we discussed their experience creating online experiments and interacting with participants. They also imagined possible challenges of including science communication pages in their experiments.

4.2 Analysis

We developed a code book where our research questions guided the initial set of high-level codes. Two authors independently open-coded one interview and discussed emergent codes. Using these preliminary codes, they re-coded the same interview and discussed problems and new codes with a third author. With the final code book, two authors coded all interviews (one author without knowledge of the research questions). The team discussed any discrepancies until reaching consensus. Inter-rater reliability was not calculated as it is rarely done with semi-structured interview data due to the possibility of applying the same code to different sections of the interview [1]. All authors conducted a thematic analysis [8] of the interviews and used affinity diagramming to uncover themes.

In the following, we refer to the researchers we interviewed as "R#(field)" (instead of the commonly used "P#") to avoid confusion with participants in the researchers' online studies. For more context, we also include researchers' fields after their number. Some of the cited quotes were slightly changed and shortened to improve readability.

4.3 Key Findings

Two major groups of themes arose during our analysis: 1) *motivations and concerns* researchers have for providing participants additional research information and 2) *steps and barriers* researchers encounter while creating conclusion pages. We describe each group's themes in detail below.

4.3.1 Motivations and Concerns. Researchers are motivated to provide participants with additional information to increase scientific knowledge and literacy and find that participants provide helpful feedback through conclusion pages. However, researchers struggle to find the right balance of information that would interest participants.

Theme 1: Researchers want to promote science learning but are skeptical about participants' interests. For most researchers providing conclusion pages, participant education was an explicit aim. The form of learning that researchers hoped their participants would take away varied from knowledge about a particular domain area (content) to how scientific conclusions are made (process). R02(medicine), a researcher who runs her own volunteer-based online experiment platform, discussed the prevalence of fraudulent online medical tests and how it was important for her online studies to counter misinformation by providing credible research-based information.

> "There's a whole lot of this crap out there. So for me it's good that it's fun and engaging and all that kind of stuff. But if there's a way that through the course of doing my research I'm providing benefit and additional resource to the universe that otherwise wouldn't be there, then that's terrific." — R02(medicine)

R09(HCI) sees conclusion pages that facilitate participant learning as crucial to rendering science more transparent and open to wider civic engagement.

> "I'd like the tests (and the [conclusion] pages) to get people to engage with scientific concepts related to the tests they just completed. I want people to understand more how the scientific process works, the fact that uncertainty in our conclusions is a healthy state of affairs, make science and scientists feel more approachable, etc." — R09(HCI)

Many of the researchers who had never provided conclusion pages had not considered the idea of sharing research information with participants previously. Once they understood what kind of information a conclusion page could communicate, they liked the idea and in hindsight thought they should have used them. For instance, R11(computer architecture), a scientist who often uses MTurk, expressed surprise that some researchers even incentivize participation with research information. He then explained what information he would have included at the end of his last experiment:

> "That sounds like something we totally should have done, explain like why we're doing this thing....I would take more or less the motivation we put in the abstract at the top of the intro...With the point being that in the future this sort of thing could show up in your daily life in interacting with computers..." — R11(computer architecture)

Even still, researchers using labor markets to conduct online experiments expressed uncertainty about the interests of their participants, assumed they were primarily motivated by monetary gain, and doubted the value of science communication pages. Although R05(NLP) acknowledged that such pages could be helpful for online communities, she was more skeptical about MTurk:

> "I'm not sure I would want to put in the effort for [developing conclusion pages] since I haven't had any feedback that indicates that additional information is something Turkers even want, since they're there for their task, get their money, and go." — R05(NLP)

Theme 2: Providing participants with research information elicits helpful feedback for researchers to improve their experiments and conclusion pages. However, researchers are concerned that providing information could compromise their privacy and bias their data.
To gather feedback from participants, researchers sometimes provide their contact information, a comment box, structured surveys, or a combination of the above. Participants use them to remark or inquire about the study. Several researchers mentioned how participants' feedback helped them to improve their studies and the additional information they provide at the end of online experiments.

For instance, R04(psychology) explained how participants helped him catch overlooked errors early and improve unclear explanations throughout the study. He discussed how participant feedback helped him assess data quality and improve participants' experiences:

> "I take feedback seriously and I look at that just to get a gauge of how people are doing. Want to see if I could bank that these are accurate responses and the other is how can I improve. " — R04(psychology)

R02(medicine) also highly valued early participants in her studies to provide feedback on the post-experiment information. She expressed frustration in her previous attempts to collect feedback from her colleagues and explained how participants gave her more valuable feedback, which she used to rapidly iterate on the studies:

> " We typically will just put stuff online, and then monitor very closely in the first 24/48 hours and see if we're getting complaints...and then iterate from there...For a new test in the first three or four days there will be a lot of changes." — R02(medicine)

Despite unparalleled feedback from participants, some researchers assumed that participants and they should keep their anonymity. R06(AI) told us an anecdote in which a MTurk participant found his professional website and emailed him to say how "really cool" his research was. Although online studies are not supposed to hide the researcher's identity, R06(AI) did not want to provide his participants more research information because "I don't want people looking me up" —R06(AI). Similarly, R07(HCI) told us that she did not want to follow up with participants outside of the data collection process:

> "It's not supposed to be double blind, but it feels sort of weird, to contact people who were supposed to be anonymous." — R07(HCI)

Furthermore, when researchers who had not provided additional information after online studies in the past were asked to think about what they might include, they repeatedly expressed concerns about additional information biasing the participants. They were concerned that participants may try to provide data they thought researchers wanted and/or tell other participants how to game the study. Researchers seemed to espouse in both words and tone a general belief that less information was better. When R11(computer science) described his study about people's visual impressions of "the tradeoffs between efficiency and quality" in images, his major hesitation for including additional information was biasing participants towards his desired outcome for the study:

> "Part of what we were trying to do was get the unbiased perspectives of someone who had never heard of our little research area...knowing too much about what they are contributing to seems like counter to the point." — R11(computer architecture)

Summary of Themes 1-2: Researchers have educational and pragmatic motivations for providing participants with research information at the end of online experiments. Researchers do not know what kind of information would interest their participants without revealing too many details about the study design or themselves. These findings suggest that strategies for identifying participant interests and then finding a middle ground between participants' interests and the researchers' experimental and privacy concerns could increase the inclusion of science communication in online experiments.

4.3.2 Process and barriers. Researchers use an iterative process to create conclusion pages, improvising ad hoc solutions to the lack of time, skills, and tools. Most researchers told us that they plan conclusion pages from the beginning of study development. Our interviews showed that researchers seek inspiration from other online experiments and non-research quizzes, develop ways to align participants' interests and their own research goals, reuse templates and pages to save time, and rely on their prior experiences and intuition for what participants appreciate.

Theme 3: Seeking inspiration takes time and pays off if numerous participants can be recruited. Most researchers reported that they initially did not know what content they should include on a new conclusion page. They frequently mentioned that they sought inspiration from other online research studies, non-research online quizzes (e.g., from *The New York Times* and *BuzzFeed*), and games. For example, R10(psychology), who created and maintains his own volunteer-based online experiment platform, told us:

> "I had seen [that a] quiz at *The New York Times* did really well the prior year. So I was inspired by that. I was also inspired by [another platform's study]. And so this seemed like something that I could predict that people might find compelling." — R10(psychology)

R01(sociology), the primary developer for a volunteer-based online experiment platform, looked for and copied engaging games by asking the following questions:

> "How do the people who are creating games reward users? What are the kinds of things that you can do in a traditional game that users find to either be fun or rewarding or in some ways incentivizing to continue to play?" — R01(sociology)

For researchers who incentivize volunteers to participate in their online studies with post-experiment information, the amount of time spent looking for popular examples and creativity exerted on a particular page is proportional to the needed sample size. For instance, R10(psychology) discussed how he had included multiple ways for people to understand their data and the research for a particular study and emphasized that even though he valued participant learning, spending time on this kind of outreach was not sufficiently valued as a contribution in his research community:

> "For this particular study I knew I needed at least half a million subjects. That's why I designed it this way. For a lot of studies, I don't need that many and so I don't necessarily bother....I'm a scientist, not a science educator. So doing a really good job at science outreach while doing bad science would count as a fail. Whereas if I did

really good science without doing any science outreach, that would be perfectly fine." —R10(psychology)

Theme 4: Gauging participant interest and providing information that appeals to them in an authoritative and understanding manner is critical to avoid misinterpretation. The tension between researchers' goals and participants' interests emerged as one of the most prominent themes in our interviews. In particular, medical and psychology researchers were concerned about how participants interpreted their results from online experiments. They imagined people were eager to take their online experiments to learn more about themselves, but, as R04(psychology), voiced, they were worried about participants misinterpreting the research information.

> "There's a recent big data project [where]... .... we were giving [information] about autistic traits...we had to be very, very careful about how we were wording it and emphasizing that the test wasn't diagnostic... you don't want anyone walking away from it discouraged or sad or ruminating about how they did." — R04(psychology)

Another related challenge researchers faced was how easily non-experts could understand the research information. Because researchers did not know how much prior knowledge participants had about their research area, they aimed to make the research information as easy to understand as possible by eliminating jargon. This process was difficult and iterative based on participant feedback, and researchers risked sounding too informal and not knowledgeable. R03(social psychology) explained how wording the experimental design information was more complicated than trying to make it as simple to understand as possible and how as a platform they decided to use more formal language with jargon:

> "...it's based on science and we want people to get a gist of the idea behind how that science works. So, that's why we explain it using that language, so that they know it's not something that we just made up. That it's based on a validated psychological instrument." — R03(social psychology)

Theme 5: Reusing templates and pages reduces time and effort. Researchers were concerned with the time, effort, and overall opportunity cost of creating conclusion pages. When faced with the choice between spending time creating a new online study and adding or improving a conclusion page to an existing study, researchers prioritized creating new studies. R04(psychology) illustrated why he usually provided minimal information after studies with an anecdote:

> "And because the volume [of participants] was coming in quite quickly and because I wanted to include more tests, it was kind of like, well, I can spend a whole lot of time creating just one test and making it really visually appealing or I could break it down and try to add a lot of different tests and maybe it won't be as appealing. And so I chose the latter because it's more useful in terms of data and research." — R04(psychology)

For the conclusion pages researchers had the time to create, they tried to expedite the process by developing templates they could reuse with minor changes across all studies on their platforms. R03(social psychology) told us how she usually makes conclusion pages from a platform-wide template that a programmer developed in the past.

> "Well, we typically always include the same information. The only thing we would change would be study specific information, like the purpose of the study and why we had them do what they did." — R03(social psychology)

At the same time, researchers were willing to spend more time creating a page if most or all of it could be reused for future studies of a similar "class." R08, a neuroscientist who taught himself to program, articulated this sentiment when he talked about how the time spent to create the conclusion page of the first experiment of a particular class to be worth the time and effort because it would reduce the time and effort for similar studies later on. On the other hand, a study that would likely not be needed in a similar form required time and effort that did not pay off.

> “[I]t takes a much larger portion of the time if you don’t sort of have prorated it across lots of different experiments of the same class.... We try and write the experiments and the code to be sufficiently general. So if we have a similar idea later, we can slot that in into the same code and save ourselves some time building these things.” —R08(neuroscience)

Theme 6: Relying on intuition gained from trial-and-error is best when tools and skills are lacking. Researchers were rarely able to mimic their inspirations exactly because they wanted to emulate studies that were in research areas outside of their own and had limited time and resources. Without time or access to programmers, researchers had to prioritize and adapt their ideas to their intuitions of what participants might perceive as informative based on past experiences.

R04(psychology) explained his ideal vision for his experiments if he had more financial resources and programming skills:

> “If I had someone who was a machine learner or into AI and they were working on this project, what I’d have them do is to constantly create a mechanism and a platform and a tool where the results and the means are constantly being updated so that you have some type of visual graph and you could see where people of different cultures score and then where your score compares to theirs. And it’s constantly taking into account all of these different factors. ” — R04(psychology)

However, given limited resources and skills, R04(psychology) relied on templates from Qualtrics that he improved slowly over time with participant feedback.

R02(medicine) described the process of learning what kind of research information participants respond positively to as a trial-and-error process. She usually does “a lot of trying things out, having people complain, try something else, people complain...” —R02(medicine). While she perceived this process as time-consuming, she found that it ultimately paid off. Over several years of running her platform, she had developed a sense for what research information participants appreciate and how to phrase this content to appeal to a diverse audience.

Others also talked about how they had developed an internal set of best practices but still wished they could rely on guidelines for creating informative pages for participants.

Researchers also offered ideas for GUI-based systems or wizards that abstract the programming details involved as solutions to delegate work to others, create online studies more rapidly, and enforce some platform standardization. In fact, one researcher has taken it upon himself and his team to develop a new framework to manage the programming details, especially the back-end.

Summary of Themes 3-6: The main barriers researchers face when iteratively creating their conclusion pages are uncertainty about participants’ interest and background knowledge about the research domain, their limited time, and the lack of appropriate tools to expedite the creative process. These barriers echo the concerns we found researchers had even if they were motivated to provide research information to participants (Themes 1 and 2). Furthermore, our findings are consistent with Law et al.’s finding that researchers are deterred from crowdsourcing because current tools require precisely defined goals and parameters even though research often has ill-defined and uncertain stages [37]. Tools that are designed for uncertainty are needed to encourage researchers to use large-scale online experiments effectively. Conclusion pages can help researchers embrace the open-ended nature of research because they can collect participant feedback, which often leads to changes in the online studies and conclusion pages themselves.

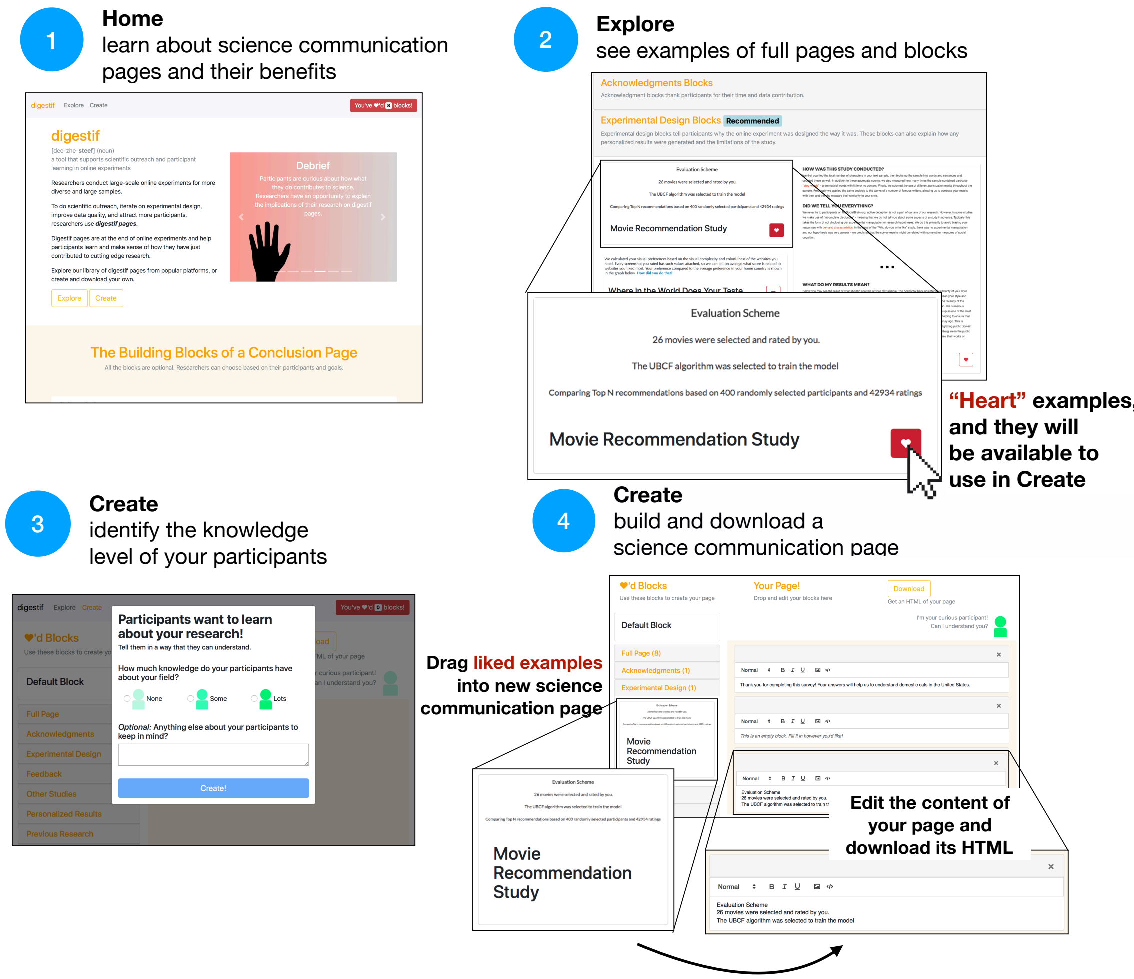

Fig. 2. Overview of Digestif. A step-by-step overview of how a researcher might use Digestif.

5 DIGESTIF AND DESIGN REQUIREMENTS

Below, we summarize findings from the two formative studies and extend them into design requirements. We instantiate the design requirements in Digestif, a web-based tool to help researchers create science communication pages for their online experiments. The target users are researchers who want to create science communication pages for the first time. Digestif is designed to appeal to researchers with a broad range of programming abilities. Digestif includes a research-based collection of reasons to use science communication pages, a library of pages for inspiration, and a canvas to create new science communication pages.

The Digestif web application was built with Python 3, Javascript, and CSS using the Django 2.11 framework [16] and a MySQL backend. Digestif can be found at http://digestif.labinthewild.org.

For each design strategy, we describe how Digestif implements it. When applicable, we also discuss alternatives we hope future systems will realize.

Allow for rapid exploration of science communication pages: During interviews, researchers told us that other studies helped them to decide how and what research information they wanted to share with participants. However, seeking inspiration and adapting existing pages into their own

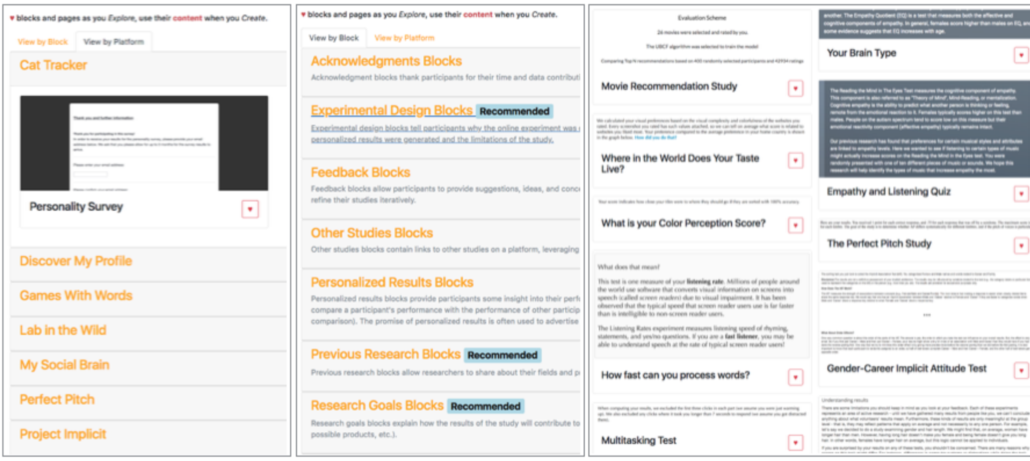

Fig. 3. Digestif's library of sample content. Left to right: 1) View of Digestif's collection of science communication pages organized by platform. 2) Recommendations of what participants appreciate based on previous research. 3) Tiling of content samples for experimental design information that researchers can "heart."

studies currently take researchers a lot of time (Theme 3). To address this problem, a system should support rapid exploration and, ideally, the adaptation of other science communication pages.

Implementation: Digestif includes a user-contributed library of pages that provide participants with research context for the online studies they take. To bootstrap the library, we included the conclusion pages we found in our landscape analysis. This content is organized by page content and platforms where the pages were used (Figure 3). The content view organizes fragments of pages based on the kind of information researchers can convey to their participants. The platform view allows researchers to peruse pages from other studies and platforms.

Support reusability of science communication pages: In our landscape analysis we found that some platforms had very similar conclusion pages across all studies. Our interviews showed us that this was likely the case because researchers reused text from pages and templates to minimize the amount of time each page required (Theme 5). Often, researchers reuse information about their long-term research goals across several studies with little to no variation because a significant portion of the research information is the same. Systems should allow for repetitive use of information so that researchers can focus on study-specific information that must change.

Implementation: A common practice among researchers creating conclusion pages was the borrowing and re-working of content that inspired them to save time and effort (Theme 5). Digestif allows researchers to create a personalized library of favorite pages and reuse the text and HTML code from the pages to make their own science communication pages (see steps 2 and 4 in Figure 2.) Researchers can also start their pages from scratch with a simplified WYSIWYG HTML editor. We chose to focus on reusability of text and minimal HTML code because a) we found in our formative studies that researchers were more concerned about the content of the page than its presentation, b) there are no empirically supported design guidelines for science communication pages, and c) we recognize that researchers may have their own styling preferences.

Surface participants' interests: In our interviews, researchers commented that they did not know what information participants would find interesting or relatable (Themes 1 and 4). As a result, researchers often relied on their intuition gained from past experiences (Theme 6). When possible, researchers also sought and incorporated participant feedback to improve the information they provide (Theme 2). Tools should recommend information participants are likely to find interesting based on literature and/or participant feedback.

Implementation: To surface participants' interests, we incorporated the following three nudges into the design of Digestif:

(1) *Recommendations*: To help researchers quickly identify information participants are likely to appreciate and find interesting, some forms of content in Digestif's library are tagged as "Recommended." The tags are based on prior work [32, 42] and intended to nudge researchers towards including the information in their science communication pages. An explanation for each recommendation is also provided when researchers hover over them, such as "In previous research, participants have liked and learned from this information."

(2) *Prompt*: Digestif also prompts researchers to answer two questions about their participants' expertise and expected interests before getting a canvas to create their science communication pages. The prompt is shown in step 3 of Figure 2. Only the first question ("How much knowledge do your participants have about your field?") is required while the second one ("Anything else about your participants to keep in mind?") is optional. The questions are intended to provoke researchers to think about writing for their participants' expertise level, which can vary depending on researchers' recruitment criteria. We intentionally asked about research "field" rather than a particular research question to encourage researchers to start by presenting a larger view of their area in which they can situate more specific questions and results.

(3) *Pictorial representation*: Once researchers answer the prompt, Digestif places a pictorial representation of a participant in the upper right-hand corner of the canvas that remains throughout the creation session. Steps 3 and 4 in Figure 2 show the prompt and resulting pictorial representation. The intention is to have a persistent participant presence as a nudge. The nudge says, "I'm your curious participant! Can I understand you?". Upon hovering, the nudge says, "Remember! I have [NO, ONLY SOME, or LOTS OF] knowledge about your research. I still want to understand your page!" depending on how much knowledge researchers expect their participants to have. If researchers also answer the optional question in the prompt, their responses are listed to the left of the pictorial participant.

Incorporate participant feedback: In our interviews, we found that researchers relied on open-ended feedback from participants who have just completed an online study to suggest improvements and share frustrations (Themes 2 and 6). In our landscape analysis we additionally found that researchers collect participant feedback by including open-ended comment boxes and their email addresses. A mechanism to collect participant feedback is integral to researchers iteratively tailoring their science communication pages to participants' curiosity.

Implementation: Digestif nudges researchers to check that they have included a place for participants to leave feedback in the form of a frequently checked email address or comment box before they can proceed to download their science communication pages.

Support researchers' privacy concerns: In our landscape analysis, we found examples of researchers who preferred different degrees of anonymity. Some remained anonymous by providing generic comment boxes while others revealed their identities and provided their personal email addresses for feedback. In our interviews, we also found that researchers were concerned about their professional privacy (Theme 2). The extent to which researchers should be public intellectuals is an ongoing conversation [5, 23]. To require that systems and platforms impose a norm of privacy or publicity is not an apolitical decision. Therefore, systems designers should carefully think about the values that are embedded in the design, and researchers should consider their positions in society, their obligations, their own ethical codes, and participant expectations when choosing what information and how much to share with participants. (See [42] for a discussion of feedback mechanisms with varying degrees of researcher anonymity.) Systems can support researchers in this process by guiding them to think through these issues with information and prompts to reflect.

Implementation: Digestif provides researchers with information and examples to make informed decisions. On the home page (step 1 in Figure 2), Digestif provides researchers with information on the pros and cons of including identifying information on their science communication pages. Digestif's library also has examples of ways to elicit participant feedback with or without sharing their direct professional contact information, such as by giving a professional email address or a general text box, respectively.

Show how researchers can avoid biasing their data: Our interviews showed that researchers were concerned about biasing participants by providing additional research and experiment information (Theme 2). Systems should therefore help researchers identify (1) when they may be sharing too much information that could bias their data and/or (2) show examples of research information (e.g., research goals, experimental design, etc.) researchers can give with varying degrees of detail.

Implementation: Digestif gives researchers examples of science communication pages with varying degrees of details in the library. Researchers can see examples of how much information they feel they can provide without compromising the integrity of their study. We chose this approach because researchers in our interviews often mentioned using other pages as a reference to determine how many details to reveal to participants (Theme 3).

Appeal to a wide range of programming abilities: The majority of the conclusion pages in our sample came from disciplines where programming is not a core competency. From our interviews, we also learned that many researchers using online experiments do not have programming backgrounds. Instead of programming, researchers use survey tools, such as Qualtrics, or reuse templates other researchers or programmers have already implemented to create their studies or science communication pages (Theme 5). If researchers want to try something outside predefined templates for their science communication pages, they are restricted by tools that require programming (Theme 6). Tools to support science communication must appeal to a wide range of programming skills and enable researchers to extend the pages based on their own needs.

Implementation: Inspired by researchers' ideas about wizards and tools such as Qualtrics, we chose to use a familiar block metaphor that researchers can use to build science communication pages using drag and drop (see step 4 in Figure 2 and Figure 3). More experienced programmers are able to download the content-focused HTML code and then augment it using their favorite web development libraries.

Minimize the time required to create a science communication page: Our interviews revealed that researchers are pressed for time and face a tradeoff between investing time in creating new studies or creating science communication pages for existing studies (Theme 5). The time required to create science communication pages should be sufficiently minimal that most researchers will not see it as a cost.

Implementation: Digestif supports creating a science communication page from exploration of inspiring content to use and reuse of pages. The consolidation of a multi-step process in one tool prevents researchers from having to collect and recreate themselves other studies' pages – which would likely require them to participate in several online studies and extract relevant content from different sources. The drag and drop functionality additionally enables fast assembly of existing blocks from multiple different studies.

6 EVALUATION

We evaluated Digestif through a between-subjects lab study with researchers who had never given additional research information to their participants. An in-lab study suited our purpose of observing how researchers use and respond to Digestif to create their first science communication pages.

We had three main evaluation questions:

(1) Can Digestif help researchers see online experiments from their participants' perspectives?
(2) Does Digestif increase researchers' inclinations to include science communication pages in their future online experiments?
(3) What are Digestif's strengths and weaknesses? How could researchers imagine using Digestif, if at all?

6.1 Participants

We recruited 23 researchers who had experience with behavioral, cognitive, or perceptual online experiments but had never provided participants with science communication pages. None of the researchers in our evaluation had participated in our interview formative study. We used snowball recruitment via word-of-mouth and email list postings. The researchers came from diverse backgrounds, including design, natural language processing, and human-computer interaction. Researchers had experience conducting online experiments using MTurk, CrowdFlower, Upwork, social networks, and other population-specific platforms and email lists (e.g., for disaster relief operators, people with specific illnesses, etc.).

6.2 Procedure

During the evaluation, participants created a science communication page for an online experiment, completed pre- and post-task surveys, and had a brief informal interview. To help researchers remember details of their online experiments, they were asked to bring notes for a recent or current online experiment.

Fourteen researchers were randomly assigned to the Digestif condition and nine to the no Digestif condition. Sixteen researchers (10 Digestif, 6 no Digestif) had conducted between 1-5 online studies, and seven (4 Digestif, 3 no Digestif) had conducted between 6-10 studies. Additionally, eight researchers (4 Digestif, 4 no Digestif) had conducted no lab studies before. This range of experience was similar to our interviewees who had not previously provided science communication pages.

In the Digestif condition, researchers were asked to use Digestif, their notes, and nothing else to create a science communication page. In the no Digestif condition, researchers were asked to create a science communication page using their notes as well as any tools and resources (except Digestif) they had or could find online. All researchers were asked to create the pages as quickly as possible without compromising quality. In the task instructions, we provided guidelines for optional information researchers could consider including in their science communication pages, which we recommended be approximately 300 words long. They were given 40 minutes to complete the task.

The pre- and post-task surveys, distributed to researchers in both conditions, asked about their experience with online studies and science communication pages, perspectives and experiences with science communication, and their knowledge about participants. The surveys were comprised of 10-point phrase completion ratings [27] and open-ended, free response questions. The pre-study survey provided baseline measures to see how researchers' experiences using Digestif affected their attitudes about science communication and participants.

The short, unstructured interview at the end of the session lasted approximately 5 minutes and allowed researchers to elaborate on their survey responses if they desired.

Our main questions structured our thematic analysis [8]. We developed themes and used them to code the open-ended responses participants gave in their pre- and post-task surveys and interviews.

6.3 Findings

In the Digestif condition, the median time to create a science communication page, which included the time to read the home page and explore the library if the researchers chose to, was 11 minutes and 40 seconds. In the no Digestif condition, most researchers used GoogleDocs to create science communication pages and took a median time of 14 minutes and 55 seconds, approximately three minutes longer than researchers who used Digestif.

Below, we refer to researchers who participated in our evaluation as "E#" to distinguish them from the researchers we interviewed in our second formative study.

6.3.1 Can Digestif help researchers see online experiments from their participants' perspectives? Researchers' open-ended responses in the pre- and post-surveys and informal interviews showed that researchers in both conditions were initially skeptical of participants' interests in learning more. However, unlike researchers who did not use Digestif, researchers in the Digestif condition expressed their realizations that participants were curious about the purpose of their participation.

For instance, E07, a researcher in artificial intelligence, repeatedly expressed how she "began to feel that [participants] are more of an organic part in [her] research, instead of just being third-party data collectors of annotators."

Furthermore, researchers mentioned how Digestif helped them to think about the range of information that their participants would be interested in. E11 expressed how Digestif helped "humanize" his participants by making him think "about what my research participants could get away from the study...I had to ask myself what kind of information they could be interesting in having, how much time they would have available for this, etc.". E06 also echoed this when she explained how Digestif helped her reflect on her participants' research experiences:

> "I think what has changed is that now I see a path to helping my participants gain a better understanding of the purpose of their work within my own experiments." — E06

Analyzing the themes and tones that researchers in the no Digestif condition expressed in their open-ended responses to questions asking about their perceptions of their participants' motivations, we found that they believed that participants were primarily motivated by monetary gain and remained skeptical that their participants would be willing to learn about research before and after creating science communication pages. For instance, E20, a visualization researcher, explained after creating a science communication page:

> "I could see it's important and valuable for the participants to know more about the experiment if they want to. However, I also doubt if anyone will ever take the time to read it ... (as I suspect most Turkers' main incentive is to get the money as quickly as possible)" — E20

6.3.2 Does Digestif increase researchers' inclinations to include science communication pages in their future online studies?

Researchers became more inclined to include science communication pages after using Digestif. When researchers were asked to rate their inclination to include them, researchers in the Digestif condition gave a median rating of 3 on a scale from 1 (never) to 10 (all future online experiments). After using Digestif, researchers' median ratings rose from 3 to 7. Researchers in the no Digestif condition gave a higher pre-survey median rating of 5 out of 10. In contrast, researchers' ratings in the no Digestif condition remained at a median of 5 out of 10 in the post-survey. Not only did the ratings of researchers using Digestif increase overall, but they also rose above and beyond that of the no Digestif condition.

The primary reason researchers in the Digestif condition gave for an increase in inclination was the ease of creating a page with Digestif. Researchers generally felt similarly to E11 and E09, who wrote in the post-task survey:

> "Currently, it takes time to create [science communication] pages like this. As far as I know, survey tools don't have features that could facilitate this process. I believe that if I could use a tool like [Digestif], I'd be far more likely to do it." — E11

> "...it is a good idea to inform participants about why we are doing this research and how others are going to benefit from it." — E09

In the next section, we discuss in detail *how* Digestif eased the process of creating science communication pages for researchers.

6.3.3 What are Digestif's strengths and weaknesses? How could researchers imagine using Digestif?

Strengths: Scaffolding, Examples, and Interaction. Researchers overwhelmingly found the blocks of content that Digestif utilized to organize the different kinds of information and to support the creation of new science communication pages to be the most useful. The examples in the library that researchers could directly manipulate were also popular among the researchers to give them a "basic idea" (E02) of what science communication pages are, the vast range of information that can be found on science communication pages, and inspiration to create their own – as was intended.

E07, an industry scientist who had conducted three to five online studies, further explained how the blocks helped her "to sort out what units are helpful to be included in a page." E09 also felt that the blocks provided some sort of "skeleton" to his science communication page and jokingly commented that the blocks could, "shame [him] into including information [he] normally exclude[s] for reasons of time."

Similarly, E08 found the organization of blocks by content area to the left of his canvas surprisingly useful: "I was surprised how small modules alongside can help me make a better [science communication] page."

While perusing the examples in the library, E12 "hearted" content and then dragged-and-dropped them into his canvas when creating his own science communication page. He thought the ability to take his favorite examples for the different kinds of content, find them organized by content area to the left of his canvas, and then drag-and-drop them to adapt directly was "so crafty."

Regardless of the programming abilities that the researchers had, they appreciated the WYSIWYG drag-and-drop interaction for dragging favorited content from the lefthand side of their create canvas into their new pages. E05, a researcher in natural language processing with a programming background, described the ability to drag-and-drop blocks to build a science communication page as feeling "more visual than just a word doc, and much nicer than rolling your own HTML."

When researchers in the Digestif condition were asked to rate if they would include the specific science communication pages they created with Digestif if they were to launch their online studies, they gave a median rating of six out of ten. In their open-ended responses justifying their ratings, most wanted to have more time to go through their pages to edit and add additional figures and examples and double check that they would not give away too much information. E15, who gave a rating of 5, answered: "I would want to discuss with my collaborators, and possibly come up with some more examples or figures that I create for the page."

Researchers in the no Digestif condition were most surprised at how quickly they could draft science communication pages – even though researchers were approximately 20% faster using Digestif. Researchers in the no Digestif condition gave a median rating of seven out of ten when asked if they would include the science communication pages they had just created. The primary reason for their high ratings seemed to be their awe at how quickly they could create science communication pages and the fact that they had a page to include. Researchers left open-ended remarks such as, "It was faster to get a first draft down than I would have expected" (E18) and "I can explain my research in 300 words?!" (E20). E19, who gave a rating of 10 out of 10 explained their high rating as

> "Mostly because I've already made it and am a huge proponent of open sourcing research and making others interesting in pursuing research in this area" — E19

Nonetheless, we found that researchers not using Digestif faced challenges that researchers using Digestif did not encounter. In response to an open-ended question about what they found most surprising about creating a science communication page without Digestif, E21 expressed how challenging the task was without knowing about participants' motivations and interests:

> "It was [a] little [more] difficult than I thought it might be. You want to include the information that potential participants would find useful, but not knowing who would be looking at the outreach page makes it difficult to decide what information to include and how." — E21

Additionally, E16 expressed how surprised he was to discover how some of the information he wrote for his science communication page could be reused by others:

> ""There can be many common/reusable components for online studies in the ML/NLP field, because they usually involve gathering training data anyways." — E16

It is interesting that researchers' difficulties in creating science communication pages without using Digestif were points that emerged as design requirements and for which Digestif was explicitly designed, affirming by comparison the advantage of Digestif's example-driven scaffolding.

Weaknesses: Integration. When asked if they would use Digestif to create future science communication pages, the researchers gave Digestif a median rating of 6.5 out of 10, where 10 was to use Digestif for all of their science communication pages. When asked to explain their ratings in text, researchers only had one main complaint: integration into existing workflows and tools, a common barrier for new tools. E11 expressed the tension between finding value in Digestif's scaffolding but wanting the system to be easy to integrate into his tools:

> "The suggestive structure can be useful for creating a page. At the same time, if my project is online why would i not want to include it in my source-code? I wish there is a way by which I could link the experimental platform (MTurk or otherwise) with digestif." —E11

Possible additional usages of Digestif. During the post-task survey and interviews, researchers imagined how they would use Digestif in the future. Researchers had interest in using Digestif throughout their workflow and for both pilot and final studies. E07 even imagined using Digestif to create an initial science communication page that participants could help to iteratively improve:

> "I could easily ask my pilot participants how to make the [science communication] page more interesting/engaging." — E07

Moreover, researchers found Digestif to be helpful as an organizational tool to gather information or as a "checklist" to verify they conveyed what they wanted to their participants.

Researchers also wanted to further integrate Digestif into their process of creating online studies overall. Researchers discussed using Digestif to inform participants of how their data would be used for research and then to give them feedback for how to provide higher quality data in the future, write experiment instructions, and create consent form pages.

In summary, researchers' inclinations to include science communication pages in general increased after creating pages with Digestif but remained the same without Digestif. Researchers also expressed difficulties in the process of creating science communication pages that were addressed in the design of Digestif. Therefore, we can conclude that the mere act of creating a science communication page is not enough to promote science communication in online experiments. It is likely that Digestif's explanation of design requirements for creating these pages is what convinced researchers. In particular, our findings suggest that support for surfacing participants' interests and scaffolding are especially important.

7 LIMITATIONS

A limitation of our work is that our landscape sample of conclusion pages in volunteer-based online experiments may have limited the scope of information we considered for science communication pages. There are other rich sources of science communication (e.g., blogs and podcasts) that include additional information, such as background of the researcher and how they became interested in a topic, that could interest participants.

In addition, our evaluation with Digestif was limited in its sample size and diversity. In particular, we excluded researchers who use online experiment and crowdsourcing platforms to collect labeled data, such as annotated images.

Because there are no existing tools for creating science communication pages, there was no obvious control. We compared Digestif to other tools researchers preferred, which was primarily Google Docs. Due to the familiarity of Google Docs, our evaluation is less favorable to Digestif, so our results are a more conservative estimation of the impact of Digestif on the process of creating science communication pages.

Although Digestif eased the process of creating science communication pages for researchers, we do not yet know how participants in online studies respond to pages made with Digestif. While they are similar to existing conclusion pages in online experiments with volunteers (where research has shown that they are appreciated by participants [32, 42]), it is possible that Digestif may make the science communication pages too uniform, or participants may prefer pages made without Digestif. We are excited to explore these questions in the future.

8 DISCUSSION, DESIGN IMPLICATIONS, AND FUTURE WORK

Our long-term goal is to make science communication pages a standard in online experiments. We believe that researchers have an ethical obligation to treat their online participants as they would their lab participants. Researchers should provide participants with information about the research context and open a channel of communication after study participation. In this work, we identified the barriers researchers face in providing participants information and developed design strategies to address these challenges. Our work extends prior work on the ethics of crowd work [30, 49, 50] by focusing on the challenges that researchers face. We urge platforms and systems to incorporate these design strategies. With Digestif, researchers are now more equipped to provide participants with post-experiment research information. Digestif not only elevates participants' needs and interests but also makes online experiments and crowdsourcing more approachable to researchers [37].

In this section, we integrate our findings, design decisions, and prior work to provide design implications to support more ethical online experiments.

Integrate Participant Nudges Throughout a System. A key barrier that kept researchers from providing more information to participants was their uncertainty about participants' interests. Digestif mitigates this barrier by incorporating research on participant interests and feedback throughout the system. Digestif provides recommendations for information to include on science communication pages based on what participants in prior work have appreciated and found interesting [32, 42], pictorial representations to remind researchers of the expected expertise of participants, and reminders to elicit participant feedback through comment boxes and email. These designs act as nudges [53], subtle design choices that our evaluation showed impact the way researchers behave and consider their participants. Because researchers are the focal users and participants play a secondary role in current platforms used for online experimentation (e.g., on MTurk, CrowdFlower, etc.) [46], nudges to consider participants are crucial to surfacing their needs and interests to researchers.

Furthermore, Digestif encourages researchers to collect feedback from participants by asking them if they want to include a comment box or email address before they can download their science communication pages. We chose this simple design because our formative interviews showed that researchers using volunteer-based online experiments sometimes collected feedback using open-ended comment boxes and email. A remaining challenge for Digestif is to support researchers in acting on participant feedback. As we learned in our interviews, acting on participant feedback is difficult for two main reasons: 1) participants can have conflicting views, making it difficult for researchers to know which concerns to prioritize and 2) the feedback participants leave is unstructured and often not actionable. Building on work in crowdsourcing and CSCW [25, 38], we plan to investigate ways to incorporate a crowd of volunteers to provide researchers with actionable feedback on how to improve their pages before launching with participants.

Finally, researchers in our formative studies and evaluation hoped that in the future they could prioritize individual participants' interests by telling them the kind of research information they would find most interesting, similar to how researchers would adapt their conversations with participants in the lab. In the future, we hope to uncover design strategies and ways for participants' diverse interests to be incorporated and represented in systems such as Digestif.

Address Researchers' Bias Concerns. Even after using Digestif, researchers were concerned that providing participants additional research information would bias their data. Digestif attempts to address this problem by showing researchers diverse examples where other researchers had shared more or less detail about their experimental design, for instance. However, these differences between sample pages were not visually prominent to the researcher and instead required reading through different pages.

Researchers' concern for the possibility of biasing data by sharing too many details is an opportune area to leverage human and machine intelligence. We imagine future systems could ask researchers for their experimental design and key metrics. Then, the system could leverage developments in text and style generation [17, 22] to propose sample pieces of text that are written about the research topic but exclude certain words and details.

Contextualize scaffolding. Researchers found Digestif's scaffolding and library of examples helpful and compelling. The scaffolding Digestif provided made the process of providing science information easier for researchers and increased their intention to include science communication pages in the future. It is not surprising that scaffolding is helpful as prior work has shown that structure in a task leads to improved outcomes [38, 48, 52]. However, the scaffolding Digestif provides was not enough alone. Researchers needed and liked the combination of scaffolding and examples of other people's science communication pages to create their own. This suggests that scaffolding without context is not enough to support a complex task, such as writing about research for a non-expert audience.

Digestif outperformed GoogleDocs and HTML editors because of the structure and examples it provided. We also found that the WYSIWYG interaction of Digestif appealed to people with varying degrees of programming experience. Digestif abstracts away the programming and also allows researchers to move directly from inspiration to creation. For researchers who would like to integrate Digestif into their workflow, a future version could be integrated with MTurk and Qualtrics as a browser-extension that suggests outlines and examples of research information researchers could include before they launch a study. In order to support more tailored and integrated help, Digestif would need a large corpus of science communication pages, which it can develop over time by asking researchers who use it to include their pages in the library.

Overall, our evaluation showed us the promising impact Digestif could have on researchers' inclinations to share research information with their participants and their perceptions of participants' interests. Could Digestif also lead to long-term improvements in how researchers relate with their participants? To assess Digestif's potential long-term impact, we intend to publicly deploy Digestif, integrating it in tools that researchers commonly use to create their online studies (e.g., jsPsych [31], Qualtrics, etc.) and assessing researchers' and participants' experiences with Digestif and science communication pages longitudinally. We hope that Digestif, and the design strategies it embodies, will make online experiments more equally beneficial for researchers and participants alike.

9 ACKNOWLEDGEMENTS

We thank the researchers who participated in our formative and evaluative studies for their time and involvement. We thank the reviewers for their feedback on the contributions and scope of this work. We also thank Daniel Epstein, Jonathan Bragg, Jim Chen, and Jared Roesch for multiple conversations about this paper's ideas and organization; Elena Agapie for early pilot testing; and Sandy Kaplan, Manaswi Saha, and Christina Chung for copy editing and feedback. This work was partially funded by NSF award #1651487.

REFERENCES

[1] David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. 1997. The place of inter-rater reliability in qualitative research: an empirical study. *Sociology* 31, 3 (1997), 597–606.

[2] American Psychological Association. 2017. Ethical Principles of Psychologists and Code of Conduct. (2017).

[3] Kimberly A Barchard and John Williams. 2008. Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behavior Research Methods* 40, 4 (2008), 1111–1128.

[4] Benjamin B Bederson and Alexander J Quinn. 2011. Web workers unite! addressing challenges of online laborers. In *Extended Abstracts of ACM Conference on Human Factors in Computing Systems*. ACM, 97–106.

[5] Nicholas Behm, Sherry Rankins-Robertson, and Duane Roen. 2014. The case for academics as public intellectuals. *Academe* 100, 1 (2014), 13.

[6] Adam J Berinsky, Gregory A Huber, and Gabriel S Lenz. 2012. Using Mechanical Turk as a subject recruitment tool for experimental research. *Political Analysis* 20 (2012), 351–368.

[7] Nathan Bos, Ann Zimmerman, Judith Olson, Jude Yew, Jason Yerkie, Erik Dahl, and Gary Olson. 2007. From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication* 12, 2 (2007), 652–672.

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[9] Amy Bruckman. 2014. Research ethics and HCI. In *Ways of Knowing in HCI*. Springer New York, 449–468.

[10] Amy S Bruckman, Casey Fiesler, Jeff Hancock, and Cosmin Munteanu. 2017. CSCW research ethics town hall: Working towards community norms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 113–115.

[11] Elizabeth A Buchanan. 1999. An overview of information ethics issues in a world-wide context. *Ethics and Information Technology* 1, 3 (1999), 193–201.

[12] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5.

[13] Marissa Burgermaster, Krzysztof Z Gajos, Patricia Davidson, and Lena Mamykina. 2017. The role of explanations in casual observational learning about nutrition. In *Proceedings of ACM Conference on Human Factors in Computing Systems*. ACM, 4097–4145.

[14] Terry W Burns, D John O'Connor, and Susan M Stocklmayer. 2003. Science communication: a contemporary definition. *Public Understanding of Science* 12, 2 (2003), 183–202.

[15] Matthew JC Crump, John V McDonnell, and Todd M Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one* 8, 3 (2013), e57410.

[16] Django. 2018. (2018). https://www.djangoproject.com/, last accessed April 16, 2018.

[17] Jessica Ficler and Yoav Goldberg. 2017. Controlling linguistic style aspects in neural language generation. *arXiv preprint arXiv:1707.02633* (2017).

[18] National Commission for the Protection of Human Subjects of Biome Beha Resea and Kenneth John Pres Ryan. 1978. *The Belmont report: Ethical principles and guidelines for the protection of Human Subjects of Research-the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*. US Government Printing Office.

[19] Christopher Frauenberger, Amy S Bruckman, Cosmin Munteanu, Melissa Densmore, and Jenny Waycott. 2017. Research Ethics in HCI: A town hall meeting. In *Extended Abstracts of Conference on Human Factors in Computing Systems*. ACM, 1295–1299.

[20] Snehal Neil Gaikwad, Durim Morina, Rohit Nistala, Megha Agarwal, Alison Cossette, Radhika Bhanu, Saiph Savage, Vishwajeet Narwal, Karan Rajpal, Jeff Regino, et al. 2015. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 101–102.

[21] Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review* 19, 5 (2012), 847–857.

[22] Marjan Ghazvininejad, Xing Shi, Jay Priyadarshi, and Kevin Knight. 2017. Hafez: an Interactive Poetry Generation System. *Proceedings of ACL 2017, System Demonstrations* (2017), 43–48.

[23] Ivor Goodson. 1999. The educational researcher as a public intellectual. *British Educational Research Journal* 25, 3 (1999), 277–297.

[24] Samuel D Gosling, Simine Vazire, Sanjay Srivastava, and Oliver P John. 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist* 59, 2 (2004), 93.

[25] Michael D Greenberg, Matthew W Easterday, and Elizabeth M Gerber. 2015. Critiki: A scaffolded approach to gathering design feedback from paid crowdworkers. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, 235–244.

[26] James C Hamilton. 1999. The ethics of conducting social-science research on the Internet. *The Chronicle of Higher Education* 46, 15 (1999), B6–7.

[27] David R Hodge and David F Gillespie. 2007. Phrase completion scales: a better measurement approach than Likert scales? *Journal of Social Service Research* 33, 4 (2007), 1–12.

[28] J J Horton, D G Rand, and R J Zeckhauser. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* (2011).

[29] Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. 2010. Quality management on Amazon Mechanical Turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*. ACM, New York, NY, USA, 64–67.

[30] Lilly C Irani and M Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proc. ACM Conference on Human Factors in Computing Systems*. ACM, 611–620.

[31] jsPsych. 2017. (2017). http://www.jspsych.com, last accessed August 20, 2017.

[32] Eunice Jun, Morelle Arian, and Katharina Reinecke. 2018. The potential for scientific outreach and learning in mechanical turk experiments. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM.

[33] Eunice Jun, Gary Hsieh, and Katharina Reinecke. 2017. Types of motivation affect study selection, attention, and dropouts in online experiments. *ACM Human-Computer Interaction, Computer Supported Cooperative Work and Social Computing* 1, 2 (2017), 56. https://doi.org/10.1145/3134691

[34] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1301–1318.

[35] Robert Kraut, Judith Olson, Mahzarin Banaji, Amy Bruckman, Jeffrey Cohen, and Mick Couper. 2004. Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the conduct of research on the internet. *American psychologist* 59, 2 (2004), 105.

[36] LabintheWild. 2017. (2017). http://www.labinthewild.org, last accessed September 13, 2017.

[37] Edith Law, Krzysztof Z Gajos, Andrea Wiggins, Mary L Gray, and Alex C Williams. 2017. Crowdsourcing as a Tool for Research: Implications of Uncertainty.. In *Computer Supported Cooperative Work and Social Computing*. 1544–1561.

[38] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 473–485.

[39] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (2012), 1–23.

[40] MySocialBrain. 2017. (2017). http://www.mysocialbrain.org, last accessed August 20, 2017.

[41] Social Psychology Network. 2017. (2017). http://www.socialpsychology.org, last accessed August 20, 2017.

[42] Nigini Oliveira, Eunice Jun, and Katharina Reinecke. 2017. Citizen science opportunities in volunteer-based online experiments. In *Proceedings of ACM Conference on Human Factors in Computing Systems*. ACM, 6800–6812.

[43] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making* 5, 5 (2010).

[44] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1364–1378.

[45] Ulf-Dietrich Reips. 2000. The Web experiment method: Advantages, disadvantages, and solutions. *Psychological Experiments on the Internet* (2000), 89–117.

[46] Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Extended abstracts of ACM Conference on Human factors in Computing Systems*. ACM, 2863–2872.

[47] SciStarter. 2017. (2017). http://www.scistarter.com, last accessed August 20, 2017.

[48] Priya Sharma and Michael J Hannafin. 2007. Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments* 15, 1 (2007), 27–46.

[49] M Silberman, Joel Ross, Lilly Irani, and Bill Tomlinson. 2010. Sellers' problems in human computation markets. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*. ACM, 18–21.

[50] M Six Silberman, Lilly Irani, and R Ross. 2010. Ethics and tactics of professional crowdwork. *XRDS* 17, 2 (2010), 39–43.

[51] British Psychological Society. 2007. British Psychological Society Report of the working party on conducting research on the Internet: Guidelines for ethical practice in psychological research online. (2007). http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf.

[52] Elliot Soloway, Mark Guzdial, and Kenneth E Hay. 1994. Learner-centered design: The challenge for HCI in the 21st century. *interactions* 1, 2 (1994), 36–48.

[53] Richard H Thaler and Cass R Sunstein. 1999. *Nudge: Improving decisions about health, wealth, and happiness*. HeinOnline.

[54] Joseph B Walther. 2002. Research ethics in Internet-enabled research: Human subjects issues and methodological myopia. *Ethics and Information Technology* 4, 3 (2002), 205–216.

[55] Andrea Wiggins and Kevin Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. In *Hawaii International Conference on System Sciences*. IEEE, 1–10.

[56] Madeline Wishart and Marion Kostanski. 2009. First Do No Harm: Valuing and Respecting the'Person'in Psychological Research Online. *Counselling, Psychotherapy, and Health* 5, 1 (2009), 300–328.

[57] Games With Words. 2017. (2017). http://www.gameswithwords.org, last accessed August 15, 2017.

Received April 2018; revised July 2018; accepted November 2018.