Self-E: Procedurally Guided Self-Experiments for Novice Health Hackers

Thesis submitted in partial fulfillment for the degree of Bachelor of Science in Computer Science

By Kevin Hongyi Ouyang
Thesis Advisor: Jeff Huang; Thesis Reader: Carsten Eickhoff

Brown University
Providence, RI
May, 2020

I want to express immense gratitude for the mentorship and support of Professor Jeff Huang, who has helped me experience the joys of doing exciting work with a community of excellent researchers. This work was made possible by the HCI Research Group, in particular the SleepCoacher team.

I also want to thank Nediyana Daskalova, who has been unrelentingly kind and supportive from the inception of the project.

Lastly, for their support and love, my gratitude to Zhen Yan, He Ouyang, Linda Ouyang, and Alan Chuang.

INTRODUCTION

The personalized healthcare movement seeks to improve the quality of healthcare through patient data [11]. The widespread proliferation of self-tracking technologies, which started out as a niche interest amongst technologists who wished to attain “self-knowledge through numbers,” [17] has played a key role in the emergence of personalized healthcare, particularly for patients who manage chronic conditions such as irritable bowel syndrome, migraines, diabetes, and multiple sclerosis. In the past decade, Apple [45], Samsung [44], and other companies have introduced wellness-centered self-tracking products, targeted to a wider user base with heterogeneous motivations for self-tracking. What began as a community of technologists and tracking enthusiasts has evolved to include a global consumer base of self-trackers who use wearables and software to produce everyday “small data” in service of health, wellness, and self-reflection.

Personal informatics has thus been hailed for its potential to empower patient-users in a new age of democratized, personalized healthcare [22]. In practice, empowering users through personal informatics requires thoughtful design considerations that are sensitive to agency, autonomy, and interpretability of data, highlighting the critical role of design research for personal informatics in human computer interaction [30] [27].

The diverse motivations of self-trackers have been characterized by Rooksby et al. as primarily reflection, behavior change, and/or interest in the technology itself [15]. These motivations shape individual tracking styles and approaches, which can include different trackers and devices, different “temporalities of tracking,” and varied ways of incorporating self-tracking into daily life, what Rooksby et al. has deemed “lived informatics” [15]. One aim of self-tracking can be diagnostic, which means the user (who may or may not be clinically diagnosed with a chronic health condition) is primarily or partially interested in uncovering causal relationships in their behavior and health outcomes. For those who manage a chronic condition like IBS, this means that self-tracking may be used to identify food triggers, which are often uniquely individual for reasons not exclusive to genetics, environment, and levels of access.

Self-trackers whose primary motivation is diagnostic typically try to extract these causal relationships from their data; however, this proves to be challenging without a structured experimental process: forming a hypothesis, then staging interventions, and finally performing statistical analysis and interpretation of the data [20]. This framework has been called self-experimentation, which itself has a rich history within the context of scientific discovery [3], and is one method through which a self-tracker can generate actionable knowledge that can inform how to shift behaviors [25]. However, self-experimentation requires prior knowledge about experiment design that many users do not have, otherwise users may reach results that are not meaningful [24].

Prior work in HCI on self-experimentation has investigated its potential domain-specific applications, such as in improving sleep, identifying food triggers, and managing migraines. However, evaluations of these systems have shown that users desire a general-purpose self-experimentation system that they could incorporate into their daily lives [16] [24]. Several other studies required the presence of a researcher in order to guide decision-making throughout the setup process [46], which is not a scalable model for incorporation into consumer self-tracking technologies.

To tackle these challenges, we designed, implemented, and deployed a general-purpose self-experimentation system called Self-E. Our system draws upon best practices and guidelines for successful single-case experimentation amongst novices, striking a balance between customizability and rigidity so that users could take full advantage of the structural support we provide in the system while being free to determine the goals, outcomes, and procedures for data sampling that incorporates most easily into the fabric of their lives. We evaluate this system by deploying it into the wild, and from user data we draw implications for future research that aims to make the benefits of self-experimentation accessible to a wider audience.

RELEVANT LITERATURE

Enabling Precision Medicine through Personal Informatics

Hekler et al. have outlined the need for a small data paradigm in medicine, which can be complementary to big data research strategies in order to bring about increased precision in description, prediction, and control of individual health [35]. Personalized or precision medicine (sometimes called P4 medicine: *predictive*, *preventative*, *personalized*, and *participatory*) thus has the potential to empower patients and ultimately deliver higher quality care [12]. A small data paradigm, which is marked by an emphasis on continuous, n-of-1 data collection and eschewing concerns around external validity and statistical norms across populations, allows for "understand[ing] dynamic, multi-causal, and idiosyncratically manifesting phenomena" [35] within subjects. The design and proliferation of self-tracking technologies such as wearables and mobile apps has been a key enabler in articulating and implementing this vision of the future of medicine [21].

Self-tracking and lifelogging, whether accomplished through wearables, mobile reporting, or even analog methods such as journaling, is studied within HCI under Personal Informatics. Coined by Li et al. in 2010, personal informatics (or personal analytics) investigates and prototypes systems that employ a user's own personally-generated data to accomplish a host of different user goals, such as behavior change, chronic illness management, or general wellness [6]. To better understand how users actually use trackers in their everyday lives, Rooksby et al. studied the "lived experience" of tracking, taxonomizing tracking activity into 5 distinct (but in practice often overlapping) categories: 1) *directive tracking*, driven by a goal such as losing weight, 2) *documentary tracking*, which seeks to document activities rather than alter them, 3) *diagnostic tracking*, which aims to determine causal links between individual behavior and outcomes, 4) *collecting rewards*, motivated primarily by collecting social or material rewards, and 5) *fetishised tracking*, which is distinguished by a "purer interest" in the tracking technology itself [15].

Both diagnostic and directive self-tracking present opportunities for self-experimentation system design research [25] [28]. Users who engage in diagnostic self-tracking often find that although causal inference could greatly assist in their capacity to accomplish a health goal (e.g. identify and avoid migraine triggers), the rigor and technicality needed to draw actionable conclusions remains a significant barrier to users [13], including even experienced self-trackers. Many self-trackers thus rely on data visualizations or statistical tools to draw insights that may be inaccurate or only correlational. Lee et al. found that self-experimentation can play a positive role not just in discovering what habits to implement (*diagnostic tracking*), but also in *directive tracking* by allowing users to hypothesize and test strategies to most effectively maintain those habits over time [25].

Quantified Self and Democratized Knowledge Production

The history of self-tracking and self-experimentation extends back to ancient times, but the proliferation of tracking devices in the last 30 years [21] has driven the widespread adoption of a set of uniquely contemporary data practices. The Quantified Self (QS) movement and its community members have played key roles in this popularization of self-tracking [37]. QS community members are self-tracking enthusiasts centrally motivated by the movement's goal of "self-knowledge through numbers."

Though self-experimentation is commonly practiced by the QS community, it has yet to find traction within a larger user base. Unlike typical users of self-tracking technologies, QS community members are often professional or amaetur scientists, engineers, or otherwise technically proficient. Despite the name "Quantified *Self,"* QS members are just as invested in generating knowledge that uniquely serves the individual as they are in sharing their findings and tips for doing rigorous science without the resources available when operating within scientific institutions [17] [21]. Thus, QS community members have written extensively on best practices for running successful self-experiments. These resources are freely available on the internet (reflecting an open, hacker-esque ethic) and are sometimes presented at in-person meetups [21].

Citizen science and biohacking are closely associated with the QS movement in that they each operate within a small data paradigm and, to different extents, incorporate democratized approaches to the generation of scientific knowledge [47]. Citizen science refers to scientific projects that incorporate the public's participation to various degrees of involvement, most commonly in data collection, but sometimes also in hypothesis generation and result analysis [40]. The biohacker movement (also known as DIYbio) also began through self-experimentation [33], though it remains largely understudied within HCI. Most broadly, biohacking describes experimentation with biology (which often can be but is not limited to experimenting with the subject-investigator's body) that happens entirely outside of an institutional or professional context [47]. Biohackers are typically motivated by a transhumanist vision [33]. Although the term "biohacker" may conjure images of implants and technological prosthetics (which more specifically characterize the *grinder* submovement), actual biohacker practices include behavioral modifications such as intermittent fasting to control the body's metabolic state, or limiting exposure to blue light to encourage the production of melatonin at night [34].

Importantly, although each of these movements describe a trend toward the democratization of science, biohackers and QS members are more thoroughly involved in every step of their self-experiment setup, execution, and analysis, whereas citizen science continues to rely on an implicit expert/non-expert divide to conduct research [33]. Recently, a study run by prominent QS members Gary Wolf and Azure Grant implemented a model of Participant-Lead Research (PLR) which aimed to use participatory cohort-based self-experimentation to produce generalizable health knowledge [40].

Self-experimentation Systems

A number of self-experimentation systems have been designed and evaluated within HCI research. SleepCoacher was designed to support the execution of 3-week long self-experiments for users to improve their sleep [19]. SleepBandits introduced further user agency by allowing users to select their sleep-focused self-experiment from a list, rather than having an experiment assigned as in SleepCoacher [38]. TummyTrials used self-experimentation to identify dietary triggers for patients suffering from irritable bowel syndrome [24]. Bioloop is a sleep-tracking self-experimentation system that uses an Oura Ring and Fitbit to collect user data and deliver individualized recommendations [48]. Trialist supports n-of-1 experimentation for chronic pain management by enabling collaboration between clinicians and patients to determine the treatments, frequency of data collection, and length of experiment for an individual's experiment [46].

Fewer systems have been developed for general-purpose self-experimentation. QuantifyMe is an automated self-experimentation system that allows participants to select one of four possible self-experiments to conduct [31]. Each of the experiments had four stages of treatments and lasted for 6 weeks. PACO supports users in creating their own experiments, which they can then invite other PACO users to join [43]. The data generated from these experiments are collated for the experiment creator. PACO does not innately support setting up a phase design, nor does it provide any statistical analysis of user data [43].

Single-case Experimental Design

The rise of P4 medicine has led to a revival in research on single-case experimental design [41]. Also known as *single-patient* or *n-of-1 trials*, single-case experimentation typically involves administering two (or more) different treatments to a single subject and analyzing the effects of those treatments on a dependent variable [9]. Within clinical settings, single-case experimentation has been noted for its potential to generate higher quality care through individualized, evidence-based health insights, and has seen widespread application in not just health, but also counseling psychology, rehabilitation, sports, and education [9].

Compared to conducting a between-subjects trial, n-of-1 trials provide both greater flexibility and more opportunities for creative experimentation. However, novices to single-case experimental design often require guidance and structure in order to conduct successful experiments that lead to meaningful results [23]. In clinical settings, guidance typically takes place via collaboration between patients and providers [41]. By contrast, many self-experimentation systems are marked by the absence of such a health provider or expert throughout the experimentation process and tend to restrict much decision-making and/or use procedural guidance as a stand-in for human expertise [19] [24], though human-in-the-loop designs have also been explored [46].

SYSTEM DESIGN

We introduce Self-E. Our exploration of prior work led us to identify two persistent issues amongst users of self-experimentation systems. First, challenges at various stages of self-experimentation can reduce adherence rates, particularly in unstructured evaluation environments [38] [31]. Data collection, for instance, can be hampered by variables being poorly defined or difficult to manipulate [16]. An over-burdensome system can lead to tracking fatigue, poor daily adherence, and other problems that can lead to failure to obtain a result at all [23] [13]. Second, users often require significant amounts of guidance and restriction so as to prevent poor experiment design [20], which can lead to dubious or unmeaningful results. A third theme, which presents several tensions with the prior themes, emerges from user desire for greater freedom, whether that freedom is to make minor changes to the experiment such as check-in timing or to alter the original purpose of the experiment [24]. Lastly, a consideration that arises specifically within the context of a general-purpose self-experimentation system is that the experimental design, scaffolding, and guidance should generalize to the many different types of self-experiments that users are likely to run.

Design Process

Our goals throughout the design process of Self-E were thus: to reduce user burden during experiment setup and operationalization, to provide as much as freedom and agency as possible to users to ensure flexibility of the tool, and to identify key pitfalls that users are likely to encounter and design around them. In order to tackle these challenges, we created handmade paper prototypes (Figure 2(a)) for the setup of pre-configured and custom experiment setups. We conducted in-person user tests with these prototypes with fellow HCI researchers as well as arbitrary individuals at the Brown University bookstore and the Student Center. These tests required a user to complete a simple task, such as start and configure a new experiment of their choosing. From these tests, we surfaced a few pain points. First, if the order of steps in setting up an experiment did not conform to user expectations, then users felt confused, frustrated, or believed that they were being asked to do repeat work (which they actually were not). For instance, many users approached the experimentation process with an objective in mind first ("I want to improve my productivity") rather than a behavior modification ("Meditation") first, as was presented in our prototype. Second, users often called attention to the number of steps needed to customize a self-experiment, indicating impatience. Third, more technical concepts such as "independent variable" created significant confusion. Even when we substituted out the jargon, users continued to need further clarification.

As we transitioned from a handmade paper prototype to a low-fidelity wireframe (Figure 2(b)), we re-ordered the experiment setup to mirror user expectations, streamlined the customization process from 11 active steps to 6 steps, and replaced technical jargon with simpler keywords which we further contextualized with illustrative examples. We then presented low-fidelity and high-fidelity interactive prototypes (Figure 3) to clinicians for additional feedback.

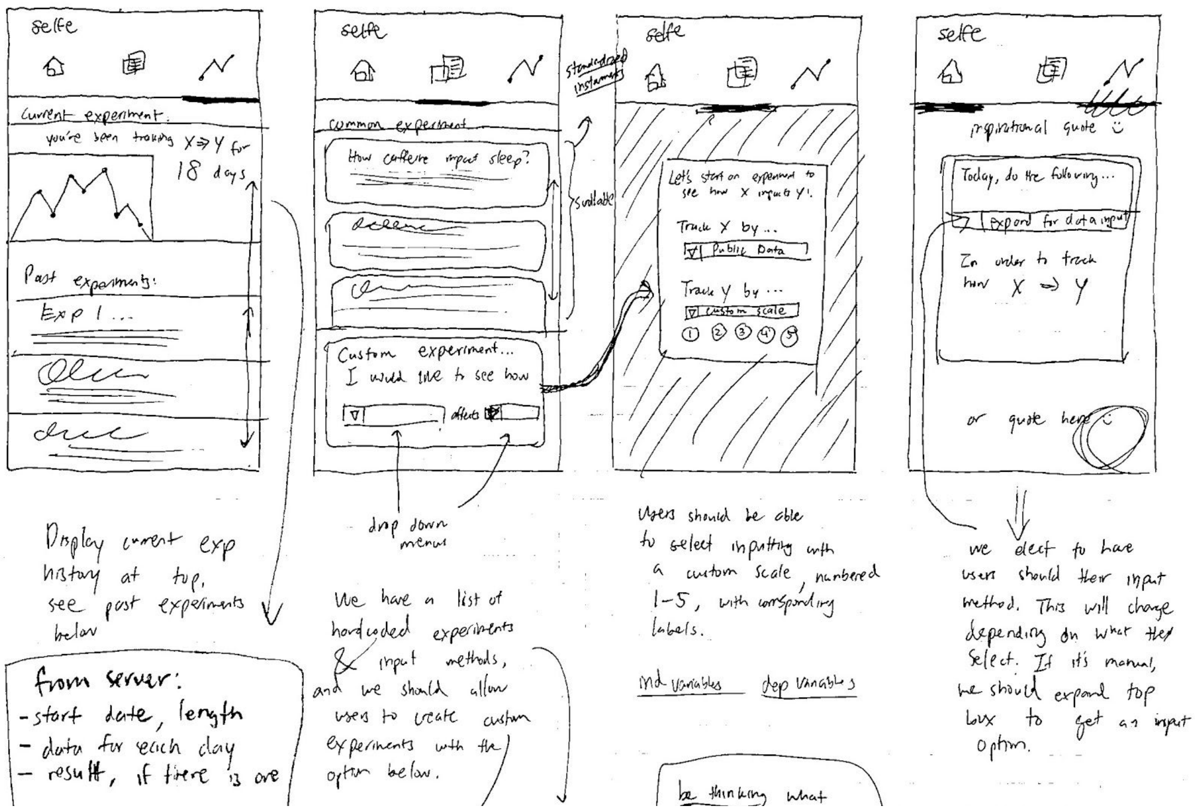


__Figure 1:__ Initial mockups for Self-E, with preliminary user flow for creating custom experiments

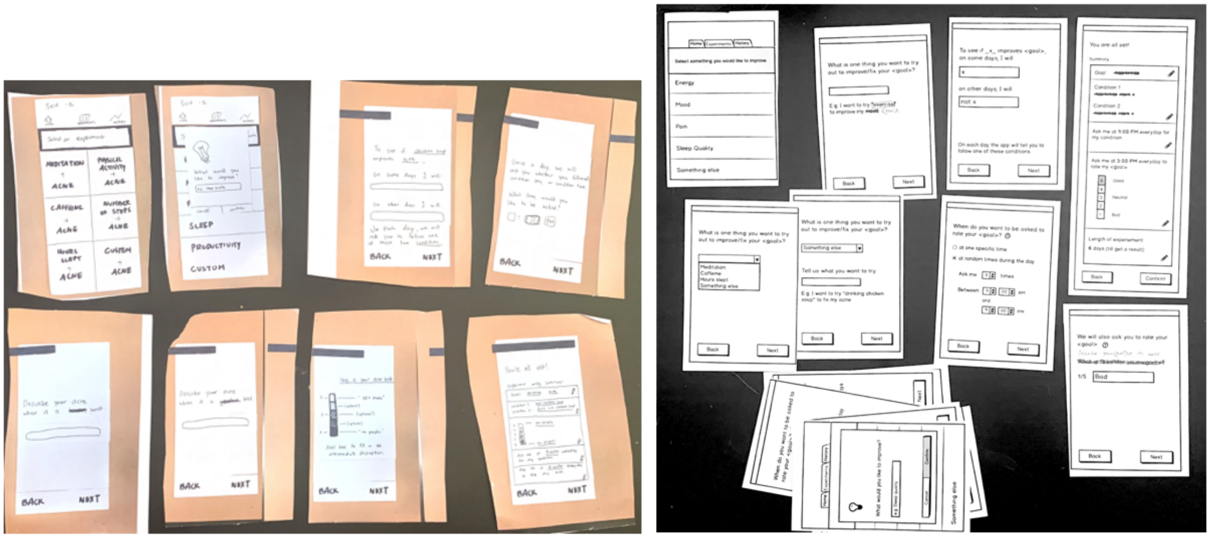


Figure 2: *(a) Handmade paper prototype used to test different user flows for customizing experiments. (b) Low fidelity prototype to further understand usability concerns.*

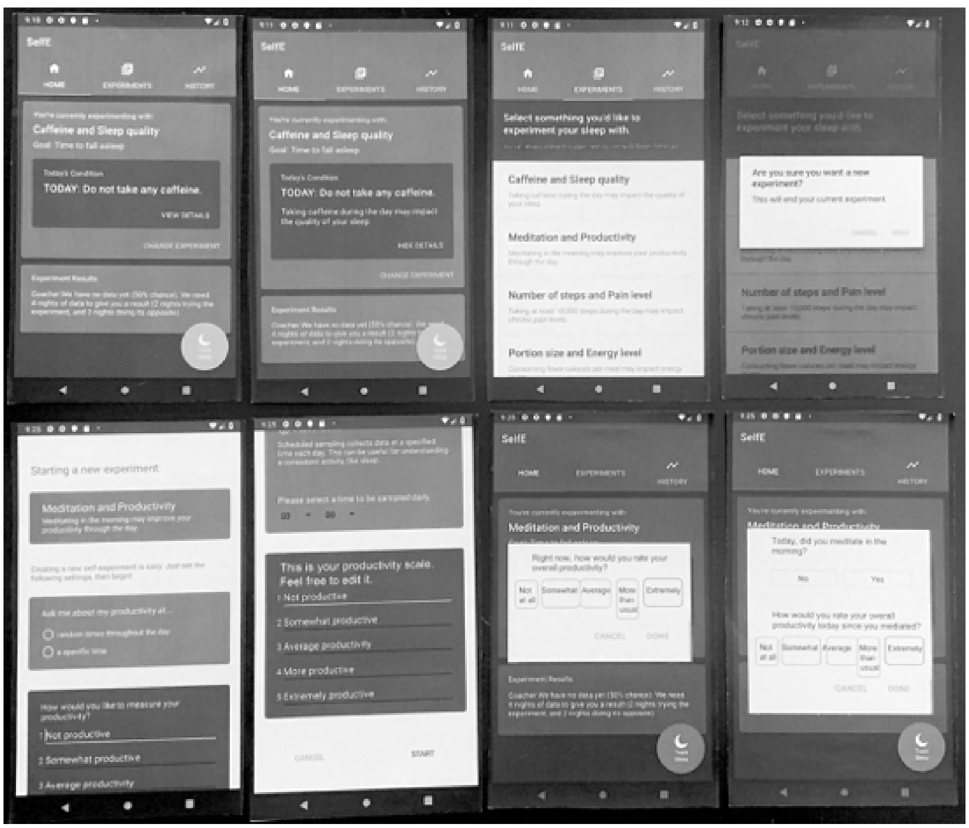


Figure 3: *High fidelity, interactive prototype.*

Flexible Experimental Design in Self-E

Building upon lessons learned from novices engaging in self-experimentation, Self-E implements a general-purpose single-case experimental design that aims to strike a balance between customizability, extensibility, and scientific rigor.

Experiments in Self-E are structured to have one independent variable and one dependent variable. To the user, these are respectively deemed the potential *cause* and the *effect* (Figure 6(a)). Users may only have one active self-experiment at a time. Barring the user from engaging in multiple experiments eliminates a source of statistical pitfalls [23]. Although many self-trackers track multiple aspects of their lives [6], and novice self-experimenters tend to be interested in running multiple experiments, prior research has demonstrated that experimenting with too many variables was a common pitfall for novices [23], leading to tracking fatigue, decreased motivation, and potentially confounding interactions between the experiments. Notably, the particular experiment structure we used in Self-E limits users in some ways. We discuss directions for expanding the flexibility of experimental design in Self-E in the discussion.

We performed a comparison of other self-experimentation platforms and their phase designs to guide our design rationale for experiments in Self-E. Previous studies have used AB [19], balanced randomized [24], and $AB_1B_2B_3$ [31] phase designs. However, more recent approaches use multi-armed bandit algorithms and Bayesian analysis to dynamically assign the more desirable condition during experimentation, which may be less disruptive to a user's daily life and present more understandable results [49] compared to balanced trials. Thus, experiments in Self-E always feature two different conditions of the independent variable which are assigned to the user daily in a randomized phase sequence weighted according to an expected reward calculated through Thompson Sampling. Being able to explore more than two levels of a condition, as implemented in QuantifyMe, may be of interest to users but requires lengthening the duration of experimenting to sufficiently explore [23]. A randomized study helps account for either known or unknown confounds [41]. In cases with carryover effect, an AB phase design may be more appropriate than a randomized one. To account for this, Self-E recommends that users only experiment with aspects that do not have a carryover effect.

Experiments in Self-E are set to be 6 days in length, but the user can alter this length for custom experiments (with guidance that still recommends a minimum of 6 days) (Figure 6(d)). Thompson Sampling can return "results" from very few data points, but premature exposure to experiment results may induce a bias in the user's recorded responses [29], so they are not exposed to the user until the number of days on which data is collected equals the set experiment length. On top of being able to provide results early on if the user so wishes, Thompson Sampling has the additional benefit of allowing an experiment to run indefinitely and adjusting a result with each new data point, giving the user great flexibility in deciding when they have gained enough insight.

Improving Quality of Self-reported Data

Self-E aims to strengthen the quality of data collected during self-experimentation by presenting the option to use the Experience Sampling Method (ESM) to gather dependent variable data (Figure 4(c)). ESM is commonly used in research that assesses affective state and technology usage [context informed scheduling] since it presents distinct advantages for measuring variables that may fluctuate throughout the day, in addition to reducing reliance on human memory by asking participants to reflect on shorter periods of time [14]. Compared with fixed scheduling, ESM provides a more holistic picture of a variable that mutates across the participant's day and works well to measure aspects such as productivity, mood, and energy level. Fixed scheduling, on the other hand, may be more appropriate for assessing aspects that do not vary throughout the day or when the goal of the experiment is to assess an aspect at a consistent time every day.

Experiments in Self-E are set to either fixed scheduling or ESM ("randomized scheduling" to the user), but users have the freedom to select either assessment protocol. If the user opts to use ESM, they may set how many times they would like to be queried throughout the day (sent in the form of app notifications), as well as a time frame within which they can field notifications. Research on mobile self-reporting has shown that the accuracy of response data may be affected by the time of day, tracking fatigue, as well as the length of questionnaire [36]. Thus, to limit user burden and encourage response accuracy, Self-E limits the time window to be at earliest 6:00 and at latest 23:45 and surveys the dependent variable using only a single 1-5 rating scale to shorten completion time. Self-E caps the number of questionnaires at 5 per day, although literature on ESM suggests that researchers use 5-17 questionnaires per day. We chose the upper limit of 5 based on literature that suggests that novice self-experimenters experience tracking fatigue even when answering fewer than 5 questionnaires per day [23].

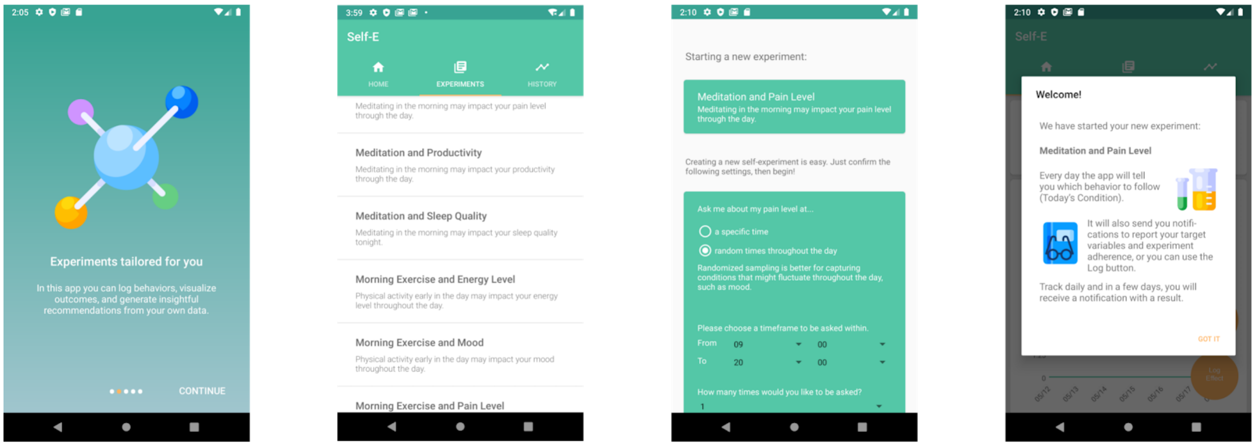
User Interface - Experiment Setup

When a new user begins using Self-E, they are taken through an on-boarding process that briefly introduces the concept of self-experimentation and its advantages (Figure 4(a)). Upon completing registration, users are required to either select an experiment from a list of pre-configured experiments or to create a custom experiment (Figure 4(b)). Pre-configured experiments are each configured by our team of researchers using recommended settings backed by research. The list of pre-configured experiments was curated to highlight the unique advantages of n-of-1 experimentation, where the user only cares about their individual results which can be impacted by complex factors such as genetics, lifestyle, and environment. For example, caffeine is the independent variable for several experiments within Self-E because research has shown that genetic factors drive differences in an individual's metabolism of caffeine, leading to disparate impacts on a person's mood and ability to sleep [4].

In order to create the list of pre-configured experiments, we drew from self-tracking literature to build a list of aspects that were commonly tracked [15], and we consulted with clinicians to generate interventions that would be both interesting for single-case experimentation and viable for our specific experimental design (e.g. unlikely to have carryover effect). The list contains 24 experiments, each of which is a combination of one of the five interventions (meditation, physical activity, food & drink, walking, and hours slept) and one of the five effect variables (energy level, mood, pain level, productivity, and sleep quality). The only combination that we did not include in the list was hours slept and pain level due to the lack of background research to support the viability of such experiments.

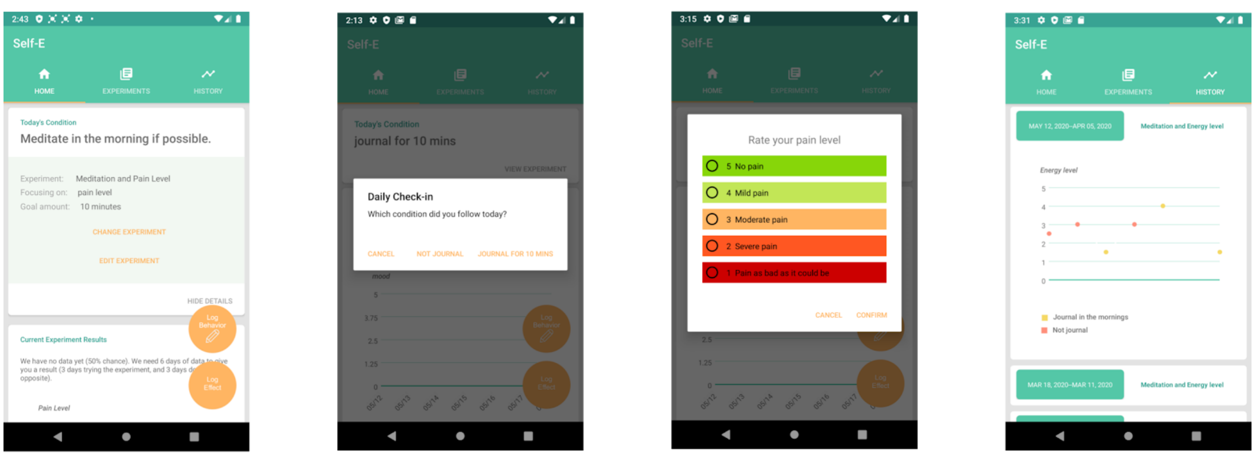
If a user selects an experiment from this list, they are taken to a settings page (Figure 4(c)). Here, they can personalize the cause variable check-in time, the effect variable check-in time window, the check-in style (fixed schedule sampling or ESM), and the amount of the intervention they will be applying (10 minutes of meditation). The flexibility that users have to change the check-in time was intended to allow them to fit self-experimentation into their schedule, to encourage a higher response rate and better adherence. Users can also customize the labels of the scale used to rate the effect variable because the ranges of experience with something like pain can vary among individuals (Figure 5(c)).

Although we present many customizable features to the user, their values are set by default with the recommendations based on prevailing research. For example, an effect variable such as "mood" can be variable throughout the day, so for experiments measuring mood, randomized experience sampling is selected by default as the check-in style [14]. The scale labels for each of the effect variables are based on a commonly accepted scale for the given variable, except for productivity which does not have a widely accepted scale, so we choose to make it general (from "very productive" to "very unproductive"). These default values alleviate friction for a novice starting an experiment for the first time, while still affording the freedom for an experienced user to alter their in-app experience to fit their needs.



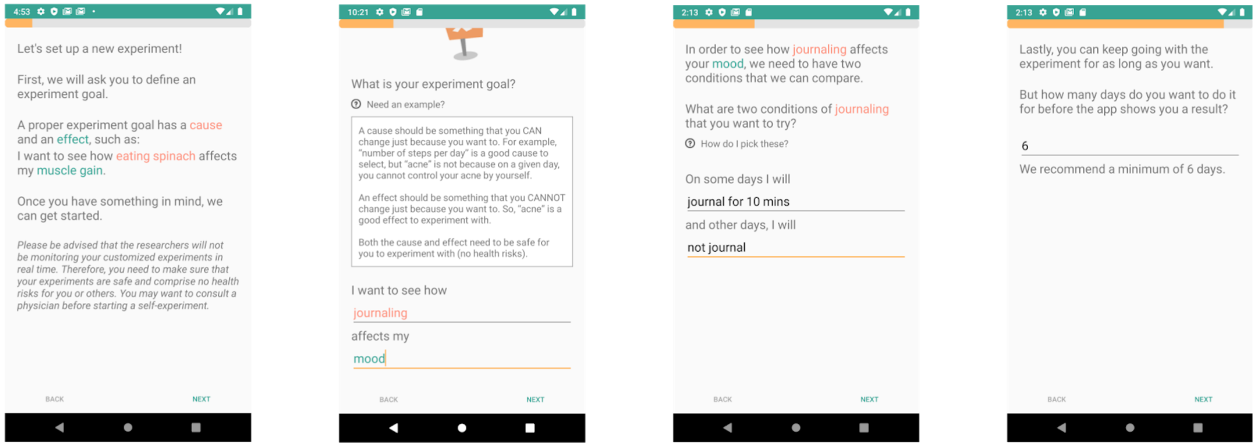
(a) (b) (c) (d)

Figure 4: *Onboarding and experiment setup. (a) Explanatory screen that introduces new users to the app. (b) Experiments tab: includes a list of pre-configured experiments as well as a custom experiment option. (c) Setting up a pre-configured experiment: fields are populated by default according to best practices, but can still be overwritten. (d) Instructions that users receive upon starting a new experiment.*



(a) (b) (c) (d)

Figure 5: *Tracking and reflecting on self-reported data during the experiment. (a) Home tab: users can self-report IV or DV data, view current experiment data or results, and edit or change their experiment. (b) Check-in popup for IV or* ***cause****. (c) Check-in popup for DV or* ***effect****. (d) History tab: users can view previous experiments that they have created.*



(a) (b) (c) (d)

Figure 6: *Custom experiment setup user flow. (a) Explanatory page for setting up a custom experiment, which uses color-coded keywords and provides an example of an experimental goal. (b) Opportunity for users to enter their custom experimental goal, along with just-in-time training on how to pick an appropriate IV and DV. (c) Condition setting: users must select two conditions (corresponding to IV) which they alternate between during their experiment. (d) Custom length: this sets the number of days that we withhold the experiment result.*

Custom Experiments

While Self-E provides a list of pre-configured experiments geared toward novice self-experimenters, users may also opt to create a fully customized experiment from scratch. In setting up a custom experiment, users can determine the independent variable, dependent variable (Figure 6(b)), two conditions to compare (Figure 6(c)), and experiment length (Figure 6(d)), on top of existing configurations that can already be made (such as rating scale labels, check-in time, etc). The custom experiments feature presents the user with an unprecedented amount of freedom in self-experimentation, allowing them to run a vast diversity of potential experiments which may not be represented by what is currently being investigated by the research community.

The freedom presented to users also presents numerous perils that may occur during setup, which may ultimately lead to poor experimental design. For example, when testing our prototypes, we found that some users created custom conditions that did not relate well to their stated experiment goal. To minimize such errors, we streamlined the process using a wizard with just-in-time guidance at key points during experiment setup (Figure 6(b)), a design strategy that other self-experimentation systems have found success with [24]. Additionally, the setup wizard uses color coding to signal connections between decisions that the user makes (Figure 6(a)). Lastly, we iterated several times on wording to remove statistical jargon in order to improve accessibility.

User Interface - Operationalization

Once the user sets up the experiment, they are taken to the home screen of the app (Figure 5(a)). Users are told to check in daily (Figure 4(d)) at a fixed time for the intervention they are tracking (we chose 8pm as it was most appropriate for our list of experiments). Check-ins are initiated via notifications sent from the server at the times set by the user. When users tap these notifications, they are taken to the app where a pop-up dialog presents either a “yes” or “no” adherence question for the intervention (Figure 5(b)) or a rating scale for the dependent variable's effect (Figure 5(c)). This data is then sent back to the server. Should a user miss the notification or decide to not answer the pop-up at that moment, they can log their data later via action buttons on the home screen (Figure 5(a)).

As a user continues to use Self-E, their data is displayed on a graph (similar to Figure 5(d)), even when the user has not tracked for the minimum duration of an experiment. Self-E requires at least three data points in each condition (e.g. “meditate” vs “don't meditate”) in order to calculate a result. This length is based on the minimum length required by the single-case experimental standards [50]. Self-E also applies an “as-treated” analysis, meaning that it only looks at which condition users actually followed on a given day, rather than whether they adhered to what they were instructed to do by the app. This type of analysis is recommended when the adherence rates are low [51]. With the understanding that users will not always consistently check their phones or go into the app, we elected to prioritize overall adherence over rigidity. This choice exemplifies the balance we attempted to achieve between experimental rigor and practicality.

User Interface - Experiment Conclusion and Reflection

Results in Self-E include a confidence interval and an effect size. Presenting a confidence interval rather than p-value has been shown to be more understandable for users not well-versed in statistics [38]. Past experiment results and data can be viewed in the History tab (Figure 5(d)), providing further opportunity for user reflection. Once results are attained, users may opt to continue tracking or to start a new experiment. Based on research that suggests running iterations of self-experiments is helpful for novices to conduct higher quality self-experimentation [23], Self-E allows users to quickly start up a new version of their current experiment from the home page (Figure 5(a)).

Implementation

The Self-E system is comprised of a backend server built in Python and a mobile client implemented in Android. User profiles are created upon registration with an email and are stored in the backend server, and any configurations or data are sent to the server throughout use of the application. Storing this data in a backend server rather than locally allows for cache-clearing without fear of losing a user's profile and history. Daily check-ins are sent to users' phones via the app notifications from the backend service. Notifications are a quick and familiar way to remind self-experimenters that they are undergoing an experiment, and many clinical trials have found success with this method of gathering data through mobile devices [36].

METHOD

To evaluate Self-E, we released the app on the Google Play Store and conducted an exploratory user study from March 1st, 2020 to May 13th, 2020. Additionally, in order to surface issues specific to the custom experiment setup process, we ran a trial with three different users through UserTesting, a platform that connects developers with user test participants who are instructed to provide detailed feedback about their experiences within the app. These two methods of exploration each provide distinct advantages: user interviews can give us a more detailed understanding of user experiences in a structured setting, while releasing the app onto the Play Store and recruiting participants through consumer channels has the advantage of being able to study real-world user behavior in a way that is minimally affected by experiment bias [38].

RESULTS

User Feedback on Experiment Customization

Due to the novelty of our custom experiment feature, we sought out more detailed user feedback through the UserTesting platform. Users were instructed to explore the app and then create a custom experiment to assess the effect of milk on their stomach pain levels. Users would be recorded on video and screen-captured for the duration of the trial. After completing the task, users responded to a 4-question written survey about their experiences with Self-E. The task we assigned to users included a hypothesis that they were willing to test: milk may affect their levels of stomach pain. As such, these tests give us a picture of how users might translate a formulated hypothesis into an experimental setup in Self-E, rather than how users might generate hypotheses on their own.

	User 1	User 2	User 3
Condition One	"Drink one glass of milk"	"Drink less than two cups of milk some days"	"Drink less than two cups of milk"
Condition Two	"Drink no milk"	"Drink two glasses of milk or more on some days"	"Drink more than two cups of milk"
IV Check-in Time	21:00	20:00	9:00

DV Check-in Style & Time(s)	ESM: 9:00AM-20:00PM	Scheduled: 20:00	Scheduled: 11:00
DV Check-in Frequency	2	1	1
Scale Labels	1=”Pain free”, 3=”Some pain”, 5=”Gutwrenching pain”	(default values) 1=”Most optimal” 3=”Neutral”, 5=”Least optimal”	(default values) 1=”Most optimal” 3=”Neutral”, 5=”Least optimal”
Duration	7	7	6

Table 1: *Inputs from users during UserTesting trial.*

According to the survey responses, none of the users found the process of setting up a custom experiment in Self-E to be frustrating. One theme that arose from two of the users was the verbosity of our instructions. One user vocalized that the task presented by the app would not have been confusing even without the additional guidance, while another user became confused only once they viewed one of the examples that we provided in hopes of achieving clarity. Specifically, this user expressed confusion at the similarity between the given example of a bad pair of conditions (“drink 1 cup of soup” some days and “drink 2 cups of soup” other days) and the given example of a good pair of conditions (“drink less than 2 cups of soup” and “drink 2 or more cups of soup”), voicing that they were too similar. The similarity of the examples, however, was intended to highlight a rather subtle idea: that the two conditions should attempt to include all possible cases of the IV.

Conversely, one user was suspicious of how simple the setup was. They “[felt] like the setup has to be a lot more complex than this.” Noting that in order to conduct effective experimentation to discover dietary triggers, they would need to account for potential confounds, adding: “You have to write down every single thing you eat. I feel like someone can mess up this experiment by eating, drinking, or doing something that they don't realize affects the experiment. I think if you want an accurate result, you want to go more specific than this." This issue is elaborated upon in the discussion section.

Overall, users indicated that they were likely to recommend the app to a friend or colleague (on a scale of 1-10 where 1=Not at all likely and 10=Very Likely, two users responded 8 and one responded 10). One user stated: "I like that they hold your hand and guide you through the whole process." Another user, unprompted by any survey question, said that they were "really curious to see the results of [the] experiment, and it was very easy to follow the steps to set up an experiment… I would definitely use this app in the future."

In-The-Wild Evaluation

Self-E was marketed as a "QS-style Self-Experimentation app" categorized under Health and Fitness. In postings on Reddit and Quantified Self forums, we indicated that Self-E was being used to conduct a trial by the Brown University HCI Group, and we proposed that Self-E could help users uncover knowledge to assist in their own lifestyle changes. We also emailed a listserv of 79 individuals who had previously downloaded SleepCoacher, another Android self-experimentation app. From these emails, 5 downloaded the app and created accounts with Self-E (a conversion rate of 6.33%).

Upon downloading the app, users were asked to sign an informed consent form if they wished to continue as participants of the study. The protocols for this study and its advertising materials were reviewed by Brown's Human Subjects Office. Users were not compensated and were free to use or delete Self-E whenever they wished. Users were also informed that they could withdraw their data from research purposes even after deleting the app.

At the conclusion of the study, the app has had a cumulative total of 147 unique installs through the Google Play Store. We limit our analysis to a group of 72 users, which excludes any accounts created prior to the beginning of the study (March 1st, 2020), any accounts associated with our research lab or acquaintances of lab members, and any accounts that were created for other studies about Self-E. Of the 72 users in this exploratory study, 15 successfully used Self-E (meaning they set up a self-experiment and collected data) for more than one day, indicating a dropoff rate of 79%. In total, these users conducted 54 days of experimentation. Each user logged their data for an average of 3.6 days. Notably, we found that out of all users that downloaded Self-E during this study, none recorded enough data to reach a result. We explore this further in the discussion.

Motivations for Self-Experimentation

As self-trackers use lifelogging apps to achieve diverse (and often multiple) goals, there may exist a breadth of reasons for users to be drawn to a tool like Self-E. Motivations for users to seek out self-experimentation are not as well understood. Previous studies have done surveys to gauge what kind of self-experiments would be interesting to users [31], but these surveys were not done with users who were specifically interested in self-experimentation. User data from Self-E may begin to help paint a picture of what kinds of goals self-experimentalists in the wild tend to gravitate toward.

95% of users who started experimenting with Self-E chose a pre-configured experiment from our list rather than creating a custom experiment. The most commonly selected pre-configured experiment goal was “Productivity,” (16 users) and the most commonly selected behavioral intervention was “Meditation” (18 users).”Pain Level” was the least popular pre-configured goal (2 users), and “Number of Steps” the least popular intervention. 6.3% of users switched experiment goals throughout their use of the app. Of these users, some switched their experiment goal to something closely related to their initial one: for example, a user started an experiment to measure “Productivity” and then later started an experiment to measure “Work Output,” demonstrating a sense of iterative refinement in this user’s experimental process.

It is worth noting that although some users may approach the app with a specific goal in mind, others may have selected or created an experiment just for the purpose of trying it out. Thus, true user preferences may be explored more robustly by examining data across a longer span of time, which would reveal how user goals shift or stay constant for self-experimentalists.

Configuration Choices

Across both pre-configured and custom experiments, 42% opted to use ESM or randomized times for DV data collection (the remainder chose fixed scheduling). We found that users who used ESM tended to track their DV for more days on average (2 days longer compared to scheduled sampling). Users who opted for ESM set a check-in window with an average duration of 11.3 hours, and an average sampling rate of 2.52 times per day. Users chose ESM for assessing dependent variables such as “Mood,” and scheduled sampling was selected for experiments that gauge “Sleep Quality.” We saw that “Productivity” and “Energy level” were assessed by both styles of data collection.

Across all 70 experiments, 88% did not alter the labels of their DV rating scale. Users who opted to change aspects of the rating scale used only qualitative measures (phrases such as “average,” “unpleasant,” “can’t focus”), and none used quantitative descriptors. Users may have opted for only word descriptors for their labels because quantitative thresholds were not explicitly included in any of the examples, instructions, or pre-configured experiments, so users would have had to discover that possibility on their own.

All users who had the option to customize duration (an option only available for custom experiments) opted to diverge from the default 6. The longest duration that a user chose was 21 days, and the shortest 5 days. The average number of days a user spent tracking in a pre-configured experiment (3.6 days) was roughly the same as the number of days spent tracking in a customized experiment (3.7 days).

DISCUSSION

Limitations

This study was conducted in order to investigate how novices behave and interact with a general-purpose self-experimentation system. As this study was largely exploratory, the primary limitation we wish to acknowledge is the sample size of the study. Despite this, we want to point out that voluntary, uncompensated participant data is especially valuable for systems design research in HCI. As Bernstein et al. articulate, “most systems studies in CHI have to pay participants to come in and use research software. Any voluntary use is better than many CHI research systems will see” [5].

Yet still, a sustained marketing effort could produce a larger pool of user data down the line. General-purpose self-tracking apps and dashboards tend to be more popular with self-tracking enthusiasts compared to self-tracking novices [8], which may help explain why a general-purpose self-experimentation app may have a wider chasm to cross in order to appeal to users in the wild. Further research may find it productive to consider how best to articulate the value proposition of self-experimentation to a mainstream audience.

Below we raise points of discussion and implications for future research which emerge from the results of both the UserTesting trial and the user study.

Onboarding Novice Self-experimentalists

While a substantial churn rate for new users on mobile applications is to be expected, Self-E's is higher than average and expresses the need for stronger emphasis on the onboarding process. Although we designed and iterated with the express intent to make experiment setup and data collection highly intuitive, we did not test different strategies for onboarding or introducing the user to potentially unfamiliar concepts. Thus, novice self-experimentalists may find it especially difficult to orient themselves within such a system if they are not given adequate knowledge to connect how the different features of Self-E all fit together.

Onboarding processes typically aim to teach the user how to use the app. Due to the (presently) niche nature of self-experimentation, we propose that a robust onboarding process for a self-experimentation system should additionally accomplish the following objectives: 1) provide knowledge (or at least gesture to resources) about self-experimentation and statistical concerns in general, and 2) encourage the user to reflect on their personal habits so that they can better mold the app to fit their lives.

We contend that while some users are satisfied with having a strongly guided, opinionated tool at their disposal, other users treat such streamlining as oversimplifying and suspect, as evidenced by one of the UserTesting participants who felt like “the setup has to be a lot more complex than this.” Thus, giving users the option to receive a more thorough educational onboarding could address these feelings by clarifying what the system can offer and what the user must take responsibility for, such as confounding variables, construct validity, etc.

By encouraging self-reflexivity about a user’s habits during onboarding, users may be more likely to make configuration preferences that work well for them. Notably, setting up a self-experiment requires a high level of self-awareness on part of the user. Future work in self-experimentation systems might investigate how to automate the selection of user preferences so that users can perform less guesswork about their own tendencies. For instance, we envision an extended, multi-day onboarding process where users only track variables naturalistically. After that, upon starting an experiment, automated suggestions could be made for check-in timings, experiment duration, and other settings.

Inspire Lifestyle Optimization

A majority of users on Self-E were drawn toward lifestyle optimization goals, such as “Productivity” and “Energy Level,” as opposed to the management of health conditions such as chronic pain. Subsequent research in self-experimentation systems geared toward the mainstream thus might incorporate a greater variety of experiments that introduce users to creative ways of optimizing their health and wellness. Drawing inspiration from the biohacking community to design these experiments may be fruitful, since many such “hacks” are not typically ascribed by mainstream health advocates and might therefore be more novel to users. Any such experiments would be carefully vetted for safety with clinicians, and any requisite disclaimers should be made prominent to the user.

Examples Impact User Creativity

Despite the wide affordances that Self-E offers, users were less inclined to run a fully customized experiment, and users mostly limited their choices to ones that they had already seen represented within the app. For instance, two participants in the UserTesting trial specified conditions that were closely modeled after the pair of conditions we gave as an example. Similarly, in the user study, none of the users who changed their rating scale labels used any quantitative descriptors. This was likely due to the fact that all of the pre-configured experiments or instructional examples have qualitative measures. Additionally, all experiments with “Mood” used ESM, presumably influenced by the fact that “Mood” was presented as an example of the kind of variable that benefits from being measured randomly throughout the day. These examples were provided to convey conceptual intuitions to users, but had the adverse effect of limiting or biasing users toward a small subset of possible choices. Although we designed the system for flexibility and extensibility, further work is needed in order to understand how to encourage users to think creatively beyond the examples presented.

Directions for Future Research

In addition to the aforementioned proposals, we would like to highlight several different directions for future work. Although all of the current self-experimentation systems currently have an all-consuming focus on the individual, many QS community members spend a significant portion of their time sharing their data, experiments, and wisdom gained [17]. These exchanges allow for the proliferation of more rigorous practices and creative interventions, on top of being social. As such, subsequent work could focus more on the design of a platform on which n-of-1 experiments and "small data" are shareable, discoverable, and replicable.

Future research may also investigate how Self-E can provide scaffolding for running Participant-Led Research (PLR) studies or other citizen science projects. As demonstrated by Grant and Wolf, PLR studies can be more rewarding, more educational, and more efficient for participants compared to traditional studies [40]. PLR explicitly aims to combine the role of the researcher and participant, and this goal aligns with the approach taken by Self-E.

Experiment structure in Self-E was designed to be flexible while limiting potential user pitfalls. Future work ought to investigate how to further extend this flexibility. Allowing users to pick phase design or specify the length of a phase would accommodate a greater diversity of potential experiments within the app. Similarly, experimenting with more than two conditions of an IV, tracking multiple DVs, or running multiple, non-conflicting experiments at the same time would increase the rate at which users acquire self-knowledge. Although novices tend to run into issues, such as tracking fatigue, when running self-experiments, more experienced self-trackers may have no issue with this and may benefit greatly from this feature. Lastly, integrations with wearables and tracking devices would introduce a number of high quality, seamless, and continuous data streams that could bolster a user's self-experiments.

CONCLUSION

Eric Topol, a cardiologist and prominent digital health expert, contends that the democratization of health and medicine is one of the most significant impacts of mobile devices [22]. With that call to action in mind, we present Self-E, a general-purpose self-experimentation system implemented in Android, which we released to the Google Play store and evaluated using behavioral data from real users. Self-E was designed to investigate how automated, user-centered systems can streamline the setup and operationalization of self-experiments and provide opportunities for reflection, discovery, and behavior change. Inspired by how the benefits of self-tracking have propagated from the "radical edge" of health hackers and early adopters to the wider public, we drew insights and best practices from the QS community to tackle the problem of how to make more accessible the advantages of self-knowledge acquisition. By conducting a 10 week study, we begin to understand the viability of a system that aims to trouble the lines between investigator/subject and expert/non-expert by offering as much freedom, flexibility, and agency as possible while maintaining the need for scientific rigor. Our investigation into how to design usable and robust self-experimentation systems aims to bring closer to fruition a more participatory, democratized vision of health and science.

WORKS CITED

[1]
S. W. Strickland, “The Ideology of Self-Knowledge and the Practice of Self-Experimentation,” *Eighteenth-Century Studies*, vol. 31, no. 4, pp. 453–471, 1998.
[2]
S. Roberts, “Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight,” *Behav. Brain Sci.*, vol. 27, no. 02, Apr. 2004, doi: 10.1017/S0140525X04000068.
[3]
W. Bains, “Truly personalised medicine: Self-experimentation in medical discovery,” *Medical Hypotheses*, vol. 70, no. 4, pp. 714–718, Jan. 2008, doi: 10.1016/j.mehy.2007.08.018.
[4]
A. Yang, A. A. Palmer, and H. de Wit, “Genetics of caffeine consumption and responses to caffeine,” *Psychopharmacology*, vol. 211, no. 3, pp. 245–257, Aug. 2010, doi: 10.1007/s00213-010-1900-1.
[5]
M. S. Bernstein, M. S. Ackerman, E. H. Chi, and R. C. Miller, “The trouble with social computing systems research,” in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA ’11*, Vancouver, BC, Canada, 2011, p. 389, doi: 10.1145/1979742.1979618.
[6]
I. Li, A. Dey, J. Forlizzi, K. Höök, and Y. Medynskiy, “Personal informatics and HCI: design, theory, and social implications,” in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA ’11*, Vancouver, BC, Canada, 2011, p. 2417, doi: 10.1145/1979742.1979573.
[7]
I. Li, A. K. Dey, and J. Forlizzi, “Understanding my data, myself: supporting self-reflection with ubicomp technologies,” in *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp ’11*, Beijing, China, 2011, p. 405, doi: 10.1145/2030112.2030166.
[8]
S. Roberts, “The reception of my self-experimentation,” *Journal of Business Research*, vol. 65, no. 7, pp. 1060–1066, Jul. 2012, doi: 10.1016/j.jbusres.2011.02.014.
[9]

J. D. Smith, “Single-case experimental designs: A systematic review of published research and current standards.,” *Psychological Methods*, vol. 17, no. 4, pp. 510–550, 2012, doi: 10.1037/a0029312.

[10]

M. Swan, “Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen,” *JPM*, vol. 2, no. 3, pp. 93–118, Sep. 2012, doi: 10.3390/jpm2030093.

[11]

N. V. Chawla and D. A. Davis, “Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework,” *J GEN INTERN MED*, vol. 28, no. S3, pp. 660–665, Sep. 2013, doi: 10.1007/s11606-013-2455-8.

[12]

M. Flores, G. Glusman, K. Brogaard, N. D. Price, and L. Hood, “P4 medicine: how systems medicine will transform the healthcare sector and society,” *Per Med*, vol. 10, no. 6, pp. 565–576, 2013, doi: 10.2217/PME.13.57.

[13]

E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, “Understanding quantified-selfers’ practices in collecting and exploring personal data,” in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*, Toronto, Ontario, Canada, 2014, pp. 1143–1152, doi: 10.1145/2556288.2557372.

[14]

R. Larson and M. Csikszentmihalyi, “The Experience Sampling Method,” in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, M. Csikszentmihalyi, Ed. Dordrecht: Springer Netherlands, 2014, pp. 21–34.

[15]

J. Rooksby, M. Rost, A. Morrison, and M. C. Chalmers, “Personal tracking as lived informatics,” in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*, Toronto, Ontario, Canada, 2014, pp. 1163–1172, doi: 10.1145/2556288.2557039.

[16]

R. Karkar, J. Fogarty, J. A. Kientz, S. A. Munson, R. Vilardaga, and J. Zia, “Opportunities and challenges for self-experimentation in self-tracking,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp ’15*, Osaka, Japan, 2015, pp. 991–996, doi: 10.1145/2800835.2800949.

[17]

D. Lupton, “Lively Data, Social Fitness and Biovalue: The Intersections of Health Self-Tracking and Social Media,” *SSRN Journal*, 2015, doi: 10.2139/ssrn.2666324.

[18]

A. Ayobi, P. Marshall, and A. L. Cox, “Reflections on 5 Years of Personal Informatics: Rising Concerns and Emerging Directions,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’16*, San Jose, California, USA, 2016, pp. 2774–2781, doi: 10.1145/2851581.2892406.
[19]
N. Daskalova *et al.*, “SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST ’16*, Tokyo, Japan, 2016, pp. 347–358, doi: 10.1145/2984511.2984534.
[20]
R. Karkar *et al.*, “A framework for self-experimentation in personalized health,” *J Am Med Inform Assoc*, vol. 23, no. 3, pp. 440–448, May 2016, doi: 10.1093/jamia/ocv150.
[21]
D. Lupton, *The Quantified Self: A Sociology of Self-Tracking*. Cambridge, UK: Polity, 2016.
[22]
E. J. Topol, *The patient will see you now: the future of medicine is in your hands*. New York: Basic Books, 2016.
[23]
N. Daskalova, K. Desingh, A. Papoutsaki, D. Schulze, H. Sha, and J. Huang, “Lessons Learned from Two Cohorts of Personal Informatics Self-Experiments,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–22, Sep. 2017, doi: 10.1145/3130911.
[24]
R. Karkar *et al.*, “TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, Denver, Colorado, USA, 2017, pp. 6850–6863, doi: 10.1145/3025453.3025480.
[25]
J. Lee, E. Walker, W. Burleson, M. Kay, M. Buman, and E. B. Hekler, “Self-Experimentation for Behavior Change: Design and Formative Evaluation of Two Approaches,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, Denver, Colorado, USA, 2017, pp. 6837–6849, doi: 10.1145/3025453.3026038.
[26]
S. Pink and V. Fors, “Self-tracking and mobile media: New digital materialities,” *Mobile Media & Communication*, vol. 5, no. 3, pp. 219–238, Sep. 2017, doi: 10.1177/2050157917695578.
[27]

T. Sharon, “Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare,” *Philos. Technol.*, vol. 30, no. 1, pp. 93–121, Mar. 2017, doi: 10.1007/s13347-016-0215-5.
[28]
R. Karkar, “Designing for Diagnostic Self-Tracking,” in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp ’18*, Singapore, Singapore, 2018, pp. 516–521, doi: 10.1145/3267305.3267314.
[29]
A. Krasny-Pacini and J. Evans, “Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide,” *Annals of Physical and Rehabilitation Medicine*, vol. 61, no. 3, pp. 164–179, May 2018, doi: 10.1016/j.rehab.2017.12.002.
[30]
A. Rapp, A. Marcengo, L. Buriano, G. Ruffo, M. Lai, and F. Cena, “Designing a personal informatics system for users without experience in self-tracking: a case study,” *Behaviour & Information Technology*, vol. 37, no. 4, pp. 335–366, Apr. 2018, doi: 10.1080/0144929X.2018.1436592.
[31]
A. Sano, S. Taylor, C. Ferguson, A. Mohan, and R. W. Picard, “QuantifyMe: An Automated Single-Case Experimental Design Platform,” in *Wireless Mobile Communication and Healthcare*, vol. 247, P. Perego, A. M. Rahmani, and N. TaheriNejad, Eds. Cham: Springer International Publishing, 2018, pp. 199–206.
[32]
J. Schroeder *et al.*, “Examining Self-Tracking by People with Migraine: Goals, Needs, and Opportunities in a Chronic Health Condition,” in *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS ’18*, Hong Kong, China, 2018, pp. 135–148, doi: 10.1145/3196709.3196738.
[33]
A. K. Yetisen, “Biohacking,” *Trends in Biotechnology*, vol. 36, no. 8, pp. 744–747, Aug. 2018, doi: 10.1016/j.tibtech.2018.02.011.
[34]
J. Hamblin, “7 Biohacks to Master Before Worrying About Other Biohacks,” *The Atlantic*, Mar. 13, 2019. https://www.theatlantic.com/health/archive/2019/03/top-biohacks/584584/ (accessed May 20, 2020).
[35]
E. B. Hekler, P. Klasnja, G. Chevance, N. M. Golaszewski, D. Lewis, and I. Sim, “Why we need a small data paradigm,” *BMC Med*, vol. 17, Jul. 2019, doi: 10.1186/s12916-019-1366-x.
[36]

N. van Berkel *et al.*, “Context-Informed Scheduling and Analysis: Improving Accuracy of Mobile Self-Reports,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, Glasgow, Scotland Uk, 2019, pp. 1–12, doi: 10.1145/3290605.3300281.
[37]
G. White, Z. Liang, and S. Clarke, “A Quantified-Self Framework for Exploring and Enhancing Personal Productivity,” in *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*, Dublin, Ireland, Sep. 2019, pp. 1–6, doi: 10.1109/CBMI.2019.8877475.
[38]
N. Daskalova *et al.*, “SleepBandits: Guided Flexible Self-Experiments for Sleep,” p. 13, 2020.
[39]
A. Ayobi, P. Marshall, and A. L. Cox, “Self-Experimentation and the Value of Uncertainty,” p. 4.
[40]
A. Grant and G. Wolf, “Design and Implementation of Participant-Led Research in the Quantified Self Community,” p. 31.
[41]
K. Rl *et al.*, “Design and Implementation of N-of-1 Trials: A User’s Guide,” p. 94.
[42]
J. Rooksby, C. McCallum, P. Asadzadeh, D. Buls, and M. Chalmers, “Some design challenges for self- experimentation apps,” p. 4.
[43]
“PACO App.” https://pacoapp.com/ (accessed May 20, 2020).
[44]
“Samsung Health | Apps,” *The Official Samsung Galaxy Site*. https://www.samsung.com/global/galaxy/apps/samsung-health/ (accessed May 20, 2020).
[45]
“Use the Health app on your iPhone or iPod touch,” *Apple Support*. https://support.apple.com/en-us/HT203037 (accessed May 20, 2020).
[46]
“Case study: Trialist for chronic pain,” *Open mHealth*. https://www.openmhealth.org/features/case-studies/case-study-trialist/ (accessed May 20, 2020).
[47]
E. Lagerspetz and M. Ahteensuu, *E pluribus unum: scripta in honorem Eerik Lagerspetz sexagesimum annum complentis*. 2016.
[48]

"Bioloop Sleep - Personalized Sleep Improvement." https://bioloopsleep.com (accessed May 20, 2020).
[49]
M. Rabbi, M. H. Aung, M. Zhang, and T. Choudhury, "MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, Osaka, Japan, 2015, pp. 707–718, doi: 10.1145/2750858.2805840.
[50]
T. R. Kratochwill *et al.*, "Single-Case Intervention Research Design Standards," *Remedial and Special Education*, vol. 34, no. 1, pp. 26–38, Jan. 2013, doi: 10.1177/0741932512452794.
[51]
M. A. Hernán and S. Hernández-Díaz, "Beyond the intention to treat in comparative effectiveness research," *Clin Trials*, vol. 9, no. 1, pp. 48–55, Feb. 2012, doi: 10.1177/1740774511420743.