

Design of a Large-scale Self-Experimentation Tool for Scientific Self-Explorations

Sayali S. Phatak

Arizona State University
Phoenix, AZ, USA
sayali.phatak@asu.edu

Elaine Chen

WNYC
New York City, NY, USA
echen@wnyc.org

Stephen M. Schueller

Northwestern University
Chicago, IL, USA
schueller@northwestern.edu

Richard L. Kravitz

UC Davis
Sacramento, CA, USA
rlkravitz@ucdavis.edu

Ida Sim

UC San Francisco
San Francisco, CA, USA
ida.sim@ucsf.edu

Christopher H. Schmid

Brown University
Providence, RI, USA
christopher_schmid@brown.edu

Eric B. Hekler

UC San Diego
San Diego, CA, USA
ehekler@ucsd.edu

ABSTRACT

Many behavioral interventions can improve health and wellbeing. However, behavior change is complex and multifaceted, and not all interventions work for everyone. Most evidence for popular interventions is from group-based studies, which poorly predict whether a given intervention would work for a specific individual. One way in which individuals can find out if a given intervention works for them is via self-experimentation using N-of-1 study designs. We are designing Hack Your Health, a self-experimentation tool where individuals can try popular health-promoting activities (e.g., physical activity, meditation) to test whether they improve their psychological well-being. In this work, we describe insights gained through user research and highlight design implications and challenges in self-experimentation in the context of public health interventions.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI); Interaction design;**

KEYWORDS

self-experimentation, personal informatics, self-tracking

INTRODUCTION

Many behavioral interventions exist to improve health and wellbeing. However, there is considerable heterogeneity in how different individuals respond to different interventions, i.e., one intervention does not work for everyone in the same manner [1, 5]. Most evidence for popular interventions is from group-based studies and analyses. While these insights are important for understanding the effect of the intervention overall, they can rarely be used to predict if the intervention would work for a given individual. Put differently, what might be an effective intervention on average (at the group level) may not be effective for specific individuals [8]. Unfortunately, while research and media often describe the benefits of different behavioral interventions that work on average, few highlight the variability in response that was observed, or how many individuals the intervention did not work for.

Behavior change is an adaptive problem that is complex and multifaceted, and various factors can influence whether or not an intervention will work for a given individual (e.g., age, preference, enjoyment, personality, motivation, social support, weather, app vs. in-person) [6, 11, 13, 14]. Moreover, many of these factors change over time. Furthermore, many interventions need to be enacted in a sustained manner over time for their impact to be maintained, highlighting the importance of not only helping individuals find interventions that 'work' in terms of improving health outcomes, but that they enjoy, and that fit into their ever-changing lifestyle. There is a strong need for scalable tools that enable individuals to engage with evidence-based interventions and systematically test if they actually work for them or not.

SELF-EXPERIMENTATION USING N-OF-1 DESIGNS

One way an individual can find out if a given intervention works for them, is via self-experimentation. Self-experimentation is a data-driven approach in which the experimenter conducts the experiment on oneself, using single-case or "N-of-1" experimental designs [12]. N-of-1 studies are conducted with a single individual as the sole unit of observation [3]. In such experiments, the participant is exposed to each intervention included in the study, and the effect is estimated by comparing outcomes within person under each intervention (or one intervention against baseline/usual routine). Such experiments account for the heterogeneity in treatment response, and offer a unique opportunity to personalize interventions by allowing decision-making to be made at the individual level. The focus is not on creating generalizable knowledge, but to aid self-discovery and create self-knowledge in a scientifically grounded way, beyond relying on intuition [12].

RELATED WORK

Over the years, the Quantified Self movement has gained momentum, with growing interest in self-tracking and self-discovery. However, prior work suggests that people often lack scientific rigor in their approach (data collection, experimental design, analysis) [2]. Recently, self-experimentation tools have been developed to provide the needed scientific support. For example, TummyTrials[7] helps those suffering from Irritable Bowel Syndrome to gain insights on foods that trigger their symptoms. SleepCoacher[4] guides users through experiments to test impact of sleep recommendations on sleep behavior. Lee et al.[10] designed a self-experimentation tool for behavior change that enabled users' self-creation of behavior change plans. We build on this work to create a tool that is accessible enough for the general population to empirically test if popular health-based interventions impact their own well-being.

HACK YOUR HEALTH

In partnership with WNYC (New York Public Radio), we are designing *Hack Your Health*, a tool that can support individuals in carrying out N-of-1 experiments to test if health activities that work in general (e.g., meditation, vigorous exercise, gratitude journaling) improve aspects of their own psychological well-being (e.g., perceived stress, focus, happiness). Hack Your Health automates the experimental design, data analysis, and provision of feedback (results). This paper describes the preliminary insights and design implications from user research carried out to inform the tool's design. As part of this user research, WNYC listeners (N=551) were surveyed, and a subset of the survey respondents (n=18) were interviewed. Demographic characteristics of both groups are provided in Tables 1 and 2.

User Research Survey

The purpose of the survey was to understand interventions and outcomes of interest to the WNYC audience. Participants were first provided a brief description of the tool, and asked to choose up to 3 outcomes (and positive/negative framing when relevant) they would be interested in improving. They were then asked to choose two interventions that they would like to test to compare

Table 1: My caption

	n (N = 551)	%
Gender		
Females	447	81.12
Males	89	16.15
Other	8	1.45
Unknown	7	1.27
Race		
White	466	84.57
Asian or Asian American	34	6.17
African American	15	2.72
Pacific Islander or Alaska Native	4	0.73
Other	3	0.54
Unknown	12	2.18
Education		
Advanced degree	311	56.44
Bachelor's degree	187	33.94
Some college	26	4.72
Associate's degree	17	3.09
High school/GED	6	1.09
Unknown	3	0.54
		Mean SD
Age (years)	43.36	13.23

and see which one (if any) improves those outcomes. Top choices for both are displayed in Tables 3 & 4. When asked about likelihood of participation, most indicated that they would be very likely (46%) or extremely likely (20%) to participate.

Interviews

Participants' survey responses were used to inform the exact questions asked in the semi-structured interview. They were asked about (1) meaningful life changes if the outcomes were to improve; (2) why the interventions interested them; (3) prior experience with the interventions; (4) reasons behind choosing the intervention combination; (7) why the tool interested them, if it did. Interviews were carried out over the phone, audio-recorded if participants consented and transcribed verbatim. Participants were first provided a brief description of the tool and informed of the purpose of the interview. Emerging themes were identified concurrently with the data collection process, and discussed with the entire research team. Data were then coded according to the identified themes. Themes were identified for all domains of interest.

Findings from Interviews

Motivations for participation included an interest in forming habits out of the activities they chose (n=12), and wanting accountability (n=9) to aid habit formation: "*I am thinking of this as a springboard to get into meditation again.*" (P4). Participants were also motivated to participate because of interest in contributing to science or a being part of a big project (n=5).

Table 2: Demographic characteristics of interview participants

	Profession	Gender	Age	Education	Race
1	Research Lab Manager	M	31	Bachelor’s degree	White/Pacific Islander
2	Computer-related work	M	35	Bachelor’s degree	White
3	Yoga instructor	F	57	Some college	White
4	Computer-related work	F	28	Advanced degree	Asian/Asian American
5	Public speaking/acting coach	F	46	Advanced degree	White
6	Sales	F	54	Bachelor’s degree	White
7	Social media manager	F	40	Bachelor’s degree	Mixed European/American Indian
8	University teacher	F	24	Associate’s degree	Asian/Asian American
9	Chef	F	27	Bachelor’s degree	White
10	Marketing Research	M	48	Advanced degree	White
11	Librarian	Other	32	Advanced degree	White
12	Manager	M	48	Advanced degree	African/African American
13	Entrepreneur	F	38	Bachelor’s degree	White
14	Graduate Student	F	25	Advanced degree	White
13	School teacher	F	34	Advanced degree	White
16	Pediatrician	F	40	Advanced degree	White
17	Consultant	F	48	Advanced degree	White
18	Software engineer	M	48	Advanced degree	White

Many participants were interested in interventions they had been unable to successfully adopt in the past. For example, P2, when asked about reasons for choosing meditation in spite of previous unsuccessful attempts, said, *"..having had a bunch of false starts at it, it's always at the back of my mind like, knowing that I haven't fully gone through with it."*

Reasons behind intervention choice included choosing ones they think are beneficial (n=6), have been advised to try by someone (n=3), believe they 'should' be doing (n = 7), or believe are backed by research (n=4). For example, P5, when asked about what she expects to happen if she tried gratitude journaling, said, *"Well, I know from the research that if you focus on things that you're grateful for, then you tend to be happier, one tends to be happier."* or P6, who had been wanting to begin exercising regularly but hadn't been able to, said *"...I'm not finding the motivation to do it for some reason even though it's something I want to do and I know that it makes me feel better and I know it's something I should do."*

It is worth noting that contrary to our expectations (though also, in hindsight makes sense based on current portrayals of these interventions in popular media), few participants seemed to doubt whether a given intervention would be valuable for them: *"Well, I definitely think that, you know, deep breathing would definitely, well I would hope, would reduce my stress, well that's what I've heard from other people, and that just the act of reducing stress alone would increase my happiness."* (P11)

Although we believed that comparing two interventions made sense from a methodological perspective, participants did not intuitively find value in comparing two active interventions. P15, who chose gratitude journaling and deep breathing – *"In my life I would see those activities almost happening simultaneously. So I don't know that it would be necessarily easy to compare."*

Table 3: Interventions included in survey

Intervention	N	%
Meditating for 10 minutes	214	39
Engaging in ten minutes of vigorous physical activity	179	32
Taking five minutes for deep breathing	179	32
Writing three things you're grateful for	171	31
Performing a random act of kindness	88	16
Laughing for five minutes	72	13
Replacing an unhealthy snack with nuts	41	7
Not drinking any caffeinated beverages in the afternoon and evening	39	7
Having a five minute conversation with a stranger	35	6
Not drinking any sugar-sweetened beverages	35	6
Giving a hug to three different people	21	4

Participants also chose activities with an intention to improve aspects of wellbeing other than/in addition to the ones they selected in the survey. Few were also interested in developing particular traits or skills through continued practice of the activity, rather than improving psychological wellbeing. For example, P17 wanted to perform acts of kindness in order to make generosity a *"fabric of who I am"*.

Table 4: Health outcomes included in survey

Health outcome	N	%
Energy/Fatigue	436	79
Energy	304	70
Fatigue	132	30
Focus/Distractio	386	70
Focus	282	73
Distractio	104	27
Serenity/Stress	374	68
Stress	194	52
Serenity	180	48
Happiness/Sadness	241	44
Happiness	187	78
Sadness	54	22
Pain	96	17
Bowel movement	57	10

As mentioned earlier, most participants wanted to try activities with an aim to form habits. They did not identify the benefit of testing whether an intervention was effective for them. A reason for this could be the media and research's lack of attention to the heterogeneity of response to these interventions that are thought to be beneficial in general, and people may not be aware of the existing variability.

Design Implications and Future Work

Design implications that emerged from this work include designing the tool to inform participants' interest in habit formation, tracking multiple health outcomes instead of a single outcome, simplifying the experimental design (e.g., testing only one intervention against their usual routine), and highlighting existing individual differences to improve understanding of the logic of self-experimentation. These findings highlighted discrepancies between the research team and users' conceptualization and expectations from Hack Your Health, and the utility of N-of-1 designs for behavioral interventions. For example, few participants expressed doubt about whether a given intervention would work for them, highlighting the possibility of unawareness, disinterest or discomfort with the uncertainty of intervention response. Previous work by Karkar et al. [7] also uncovered tensions between results from self-experimentation and peoples' lived experiences. They also found that participants incorrectly interpreted findings, or confused 'no evidence' of effect to mean that there was no evidence of the tested food triggering their symptoms. Such tensions may be exacerbated when it comes to popular public health interventions, which are all expected to be beneficial. Moreover, as Kravitz et al. [9] suggest, there is a chance that learning that an intervention believed to be effective doesn't work may have a negative impact.

Future steps include conducting prototype testing and a formative evaluation of Hack Your Health to examine such tensions, and gain a clearer understanding of user burden, perceived utility of n-of-1 experimentation in the context of behavioral interventions, and to examine users' conceptual understanding of self-experimentation.

REFERENCES

- [1] Claude Bouchard and Tuomo Rankinen. 2001. Individual differences in response to regular physical activity. *Medicine and science in sports and exercise* 33, 6 Suppl (2001), S446–51.
- [2] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *CHI. ACM*, 1143–1152.
- [3] Jesse Dallery, Rachel N Cassidy, and Bethany R Raiff. 2013. Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of medical Internet research* 15, 2 (2013).
- [4] Nediya Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoach: A personalized automated self-experimentation system for sleep recommendations. In *UIST. ACM*, 347–358.
- [5] Madhav Goyal, Sonal Singh, Erica MS Sibinga, Neda F Gould, Anastasia Rowland-Seymour, Ritu Sharma, Zackary Berger, Dana Sleicher, David D Maron, Hasan M Shihab, et al. 2014. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA internal medicine* 174, 3 (2014), 357–368.
- [6] Eric B Hekler, Matthew P Buman, Jennifer Otten, Cynthia M Castro, Lauren Grieco, Bess Marcus, Robert H Friedman, Melissa A Napolitano, and Abby C King. 2013. Determining who responds better to a computer-vs. human-delivered physical activity intervention: results from the community health advice by telephone (CHAT) trial. *International Journal of Behavioral Nutrition and Physical Activity* 10, 1 (2013), 109.
- [7] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. 2017. Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers. In *CHI. ACM*, 6850–6863.
- [8] Richard L Kravitz, Naihua Duan, and Joel Braslow. 2004. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly* 82, 4 (2004), 661–687.
- [9] Richard L Kravitz, Debora A Paterniti, M Cameron Hay, Saskia Subramanian, Dionne Evans Dean, Thomas Weisner, Sunita Vohra, and Naihua Duan. 2009. Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. *Contemporary clinical trials* 30, 5 (2009), 436–445.
- [10] Jisoo Lee, Erin Walker, Winslow Burleson, Matthew Kay, Matthew Buman, and Eric B Hekler. 2017. Self-experimentation for behavior change: Design and formative evaluation of two approaches. In *CHI. ACM*, 6837–6849.
- [11] Sonja Lyubomirsky and Kristin Layous. 2013. How do simple positive activities increase well-being? *Current directions in psychological science* 22, 1 (2013), 57–62.
- [12] Allen Neuringer. 1981. Self-experimentation: A call for change. *Behaviorism* 9, 1 (1981), 79–94.
- [13] Jo Salmon, Neville Owen, David Crawford, Adrian Bauman, and James F Sallis. 2003. Physical activity and sedentary behavior: a population-based study of barriers, enjoyment, and preference. *Health psychology* 22, 2 (2003), 178.
- [14] Stephen M Schueller. 2010. Preferences for positive psychology exercises. *The Journal of Positive Psychology* 5, 3 (2010), 192–203.