A Systematic Review and Meta-Analysis of Research on Goals for Behavior Change

Jun Zhu
Informatics
University of California, Irvine
Irvine, California, USA
junz22@uci.edu

Sruzan Lolla
School of Computing and Communications
Lancaster University
Lancaster, United Kingdom
s.lolla@lancaster.ac.uk

Meeshu Agnihotri
Department of Informatics
University of California Irvine
Irvine, California, USA
meeshua@uci.edu

Sahar Asgari Tappeh
Lancaster University
Lancaster, United Kingdom
s.asgaritappeh@lancaster.ac.uk

Lala Guluzade
School of Computing and Communications
Lancaster University
Lancaster, United Kingdom
guluzadl@lancaster.ac.uk

Elena Agapie
Department of Informatics
University of California, Irvine
Irvine, California, USA
eagapie@uci.edu

Corina Sas
School of Computing and Communications
Lancaster University
Lancaster, United Kingdom
c.sas@lancaster.ac.uk

Abstract

HCI research on goals and behavior change has significantly increased over the past decade. However, while emerging work has synthesized personal informatics goals, fewer efforts have focused on also integrating HCI research on behavior change to chart future research directions. We conducted a systematic review of 180 papers focused on goals and behavior change from over 10 years of SIGCHI journals and conference proceedings. We further analyzed 37 papers from the data set that included evaluations of interventions' effectiveness in-the-wild. We also reported on the effectiveness of 76 of such technology-based interventions and the meta-analysis of 28 of these interventions. We find that most research has focused on goals in the health and wellbeing domains, centered on the individual, low intrinsic goals, and partial use of theoretical constructs in technology-based interventions. We highlight opportunities for supporting multiple-domain, social, high intrinsic, and qualitative goals in HCI research for behavior change, and for more effective technology-based interventions with stronger theoretical underpinning, supporting users' awareness of deep motives for qualitative goals.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714072

CCS Concepts

• **Human-centered computing → HCI theory, concepts and models**; *Empirical studies in HCI*; • **General and reference → Surveys and overviews**.

Keywords

systematic review, meta-analysis, goals, behavior change, intervention effectiveness

ACM Reference Format:
Jun Zhu, Sruzan Lolla, Meeshu Agnihotri, Sahar Asgari Tappeh, Lala Guluzade, Elena Agapie, and Corina Sas. 2025. A Systematic Review and Meta-Analysis of Research on Goals for Behavior Change. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3706598.3714072

1 Introduction

Goals are desired states that people strive towards or are encouraged to attain [19], playing an important role in motivating behavior [25]. They support our sustained engagement and underpin a wide variety of behaviors, including work-related activities, everyday wellbeing practices, the maintenance of social relationships, or the management of chronic conditions. Goals are an important component of research on behavior change as they underlay people's intentions of striving towards desired behaviors [83, 148]. In HCI behavior change literature, there is a focus on how goals can be supported, particularly in personal informatics research [68]. Tracking technologies are overwhelmingly designed and used to support behavior change toward tracked goals [68]. Reviews of HCI literature that touch on people's goals have only focused on goals and behavior change in the context of tracking technologies [62, 68]. These reviews suggest an overwhelming focus of research on individual technologies to support people's goals through trackers [68], specific goal domains such as health or nutrition, and types of goals, such as quantitative goals which can be easily tracked (i.e., step counts) [62].

As research on HCI behavior change has grown, such reviews are essential, offering the opportunity to pause and take stock of what has been done, and to reflect on future research directions. One of the aspects less explored in past reviews is the intersection of behavior change and technologies supporting goals. The prior reviews that involve goal-related research in HCI focus on tracking literature [62, 68], rather than a broader understanding of goal-related research in behavior change in HCI. In particular, we have a limited understanding of the extent to which HCI literature evaluates the impact of goal-related technologies on actual changes in behaviors, along with the approaches used to support behavioral changes in HCI. In the past decade, HCI researchers have debated how theory can inform the design of behavioral technology [45], what types of evaluation measures and methods are appropriate for the types of prototypes that HCI researchers create; proposing efficacy and understanding user experiences as appropriate metrics [116], and what durations are appropriate for deploying and testing behavioral technology [93, 116]. Although past research [45, 93, 116] highlighted the importance of our approach to studying and designing for behavior change in HCI, we need to further understand the landscape of how HCI researchers design (e.g., theoretical underpinning), and evaluate goals in behavior change technologies (e.g., metrics used) and the effectiveness of the interventions they design (e.g., actual measures of change in behavior). We adopt the working definition of effectiveness as the beneficial outcome of a technology-based intervention evaluated in real work settings [80]. One way to address this gap while leveraging the large body of goal-related HCI research for behavior change is through efforts to synthesize it. In particular, our work complements previous reviews through its focus on understanding how goals are studied and designed for in HCI behavior change literature, and the effectiveness of technology-based interventions for behavior change. We conducted a systematic review in two steps. First, we reviewed 180 papers on goals and behavior change in HCI to understand what behavior change goals are studied and the technology space of those studies. Second, we scoped down our review to 37 papers that reported the effectiveness of a total of 76 technology-based interventions evaluated in-the-wild, and in addition, we conducted a meta-analysis of 28 of these interventions targeting step count or screen time. Thus, our review focuses on the following research questions, whose rationale is detailed in the Background section.

RQ1. What behavior change goals have been explored by HCI research, and what is the nature of those goals?

RQ2. How are goals studied by HCI research on behavior change, and with whom?

RQ3. Which technologies are studied by HCI research on behavior change goals, and how?

RQ4. What is the effectiveness of technology-based interventions for behavior change goals evaluated in-the-wild?

Our research contributions include: (1) understanding goals studied in HCI behavior change literature and the nature of those goals with regard to sociality and motivation; (2) understanding approaches used in the studies such as contributions, methods, measures, and stakeholders involved; (3) understanding technologies studied, or designed to support goals, theory used, site, and duration of deployments; and (4) understanding the effectiveness (actual change in behavior) of technologies for goal setting and behavior change (measures used for evaluating effectiveness, which interventions are effective).

Key takeaway findings highlight that:

- Goal domains were overwhelmingly focused on health, especially physical activity, followed by digital wellbeing, with limited focus on other domains like sustainability or finance.
- The studied goals tended to be individually rather than collectively pursued with others (e.g., family, friends, coworkers).
- Only a minority of studies (<25%) recruited participants who were intrinsically motivated to pursue the goal.
- Papers used a broad range of goal measures, most commonly behavioral measures which tend to be self-reported, or custom-made with limitedly reported validity.
- The stakeholders involved as participants in the studies were predominantly from the non-clinical population and less from the clinical population or vulnerable users.
- Many papers mentioned one theory, such as goal-setting or self determination, but less than half used theories for design.
- A large set of papers evaluated technologies, most often in-the-wild, the most common deployment site being everyday life, and the deployment duration being between one and four weeks.
- In the corpus of 180 papers, behavioral measures (e.g., step count, sleep duration) were the most common, followed by motivation and goal attainment measures.
- Among the 37 papers that evaluated interventions in-the-wild, the most effective interventions were goal setting and feedback for physical activity, and behavior cost for digital wellbeing.

While HCI research's current focus on goals for behavior change is strongly skewed towards single-domain, individual, low intrinsic, and quantitative goals, we highlight opportunities for advancing the design space of goal-related research for behavior change in HCI. These include extending the focus to explore multiple-domain, social, high intrinsic, and qualitative goals. We also suggest more inclusive stakeholders from domain experts to people living with diagnosed conditions, and children. Finally, we argue for deepening efforts to design more effective technology-based interventions by ensuring stronger theoretical rationale, supporting users' awareness of deep motives for qualitative goals, and goal setting for self-set goals, or social goals, and scaffolding reflection on tracked goals for behavior change.

2 Background

Goals have been studied for more than half a century [150], with theories and taxonomies that conceptualize goals and goal-related activities informed by psychology literature [38, 83, 148, 150]. Although goals have a long research history in HCI, researchers started designing tools to support goals for behavior change and discuss our field's approaches only in the late 2000s [44, 45, 161], thus growing the HCI literature on behavior change goals. Since then, there have been limited papers taking stock of how goals are studied in HCI literature. For instance, Epstein et al.'s [68] mapping review of Personal Informatics suggests an overwhelming focus on individual technologies to support goals through trackers but does not unpack how that literature studies goals.

Ekhtiar et al.'s [62] review on goals in personal tracking found a main focus on quantitative and health goals. They also identified an increased number of papers studying goals related to multiple behaviors and diversity in the characteristics of goals. They proposed the need to better facilitate goal setting by reducing users' burden and supporting their self-efficacy while navigating complexity as goals evolve over time. Their investigation of goals sheds light on personal informatics goal-setting literature but does not contextualize the landscape of research on goals in behavior change, which is expected to be much larger as our literature surfaces many more papers in a similar time period. Although prior research has shown that personal informatics goals are overwhelmingly focused on health [62], research outside HCI has indicated a much broader range of goals that people experience, including relationships, spiritual, entertainment, or creativity goals [38], which suggests a richer space for much more goal exploration than what was surfaced in personal informatics literature [62, 68].

What we mean by behavior change goals-related research: In this review, we intend to understand the HCI work that engages with behavior change research and goal-related activities, as defined in prior work studying goal phenomena, such as goal setting [148], planning [83], pursuit or implementation of goals [149]. In HCI such papers might involve explicit studies of techniques for goal-related activities, such as setting better exercise goals [137], novel planning techniques [6]; or papers that study activities done in service of supporting engagement with goals, such as supporting better portion estimation [34], or visualization techniques to inform eating healthier [59].

RQ1 Rationale. The aim of RQ1 is to unpack the different goals studied across domains, including but not limited to wellbeing and health. Because literature beyond HCI has pointed at goals being interconnected [125, 127], we looked at whether literature included goals studied in isolation (e.g., exercise goals) or as interconnected. Two main characteristics of goals are their sociality and type of motivation. Sociality highlights the importance of the social context of motivation [74], and while findings have shown the value of social support interventions [94], sociality has been less explored in HCI in terms of individual vs social goals [38]. With respect to the type of motivation, intrinsic (i.e., one's own interests) motivation has been shown to be more effective than merely external motivation (i.e., external reinforcement) [56, 129], albeit less explored in HCI research on goals for behavior change.

RQ2 Rationale. The aim of RQ2 is to understand how goals have been studied (e.g., contributions, approach, methods, measures) and with whom (e.g., stakeholders). While HCI contributions [259] have been previously explored in HCI systematic reviews such as the one on personal informatics [68], they have been less used to investigate the goal-related research for behavior change. Similarly, limited work has focused on the different measures of goals and their validity. The latter also aligns with researchers' call for improving the way we capture user data, arguing to expand self-reports with automatic measures [93]. There has been limited HCI work reviewing research methods and approaches within behavior change literature. Researchers have also called for being intentional and acknowledging when and for what populations we are designing and whether interventions are effective [116, 117], hence our focus on stakeholders.

RQ3 Rationale. With RQ3, we wanted to understand the technology space described in the reviewed work, their theoretical underpinning, and their evaluation. Researchers in HCI have called to action to incorporate theory into behavior change technology design [45, 93]. Yet, limited work has explored the use of various theories informing design in this space, and the breadth of technologies studied by HCI scholars interested in behavior change goals.

RQ4 Rationale. RQ4 focuses on a subset of papers describing technology-based interventions for behavior change deployed in the wild and their effectiveness [80] and the most promising technologies and goal measures supporting them. With respect to types of interventions, Ekhtiar et al. [62] identified the most common technology-based interventions in personal informatics as tracking and reflecting on data at the expense of preparation (i.e., deciding what and how to track) and acting on the insights gained from tracked data. However, we know little about the broader range of intervention types [171] studied in behavior change goals literature, their effectiveness, and how to design the most effective interventions. This RQ also aligns with HCI scholars' call for technology implementation in real-world settings for sustained use [159, 236], and for generating evidence about what interventions are effective [116, 117].

3 Systematic Review Method

We describe the selection of corpus of papers for the systematic review (N=180), and a subset of those reporting on the effectiveness of 76 behavior change technology-based interventions deployed and evaluated in-the-wild reported in 37 papers. We also outline the coding process and the main codes.

3.1 Selection of Paper Corpus

Engagement with goals is described in behavior change psychology literature through activities of setting goals [148], planning for goals [83], and implementing or pursuing goals [148]. We chose to build our corpus by searching for terms that were related to these processes. While building the corpus by searching for papers in the ACM digital library (DL), we also identified tracking as a commonly discussed goal process. To create our corpus, we searched for papers that included "behavior change" and one of the terms (and linguistic variations) "goal + ": "setting," "planning," "pursuit," "tracking," "managing," "monitoring," "implementing." We searched the ACM DL in the period January 2012 - June 2023, published in SIGCHI-sponsored conferences and affiliated publication venues.

Inclusion and exclusion criteria for paper corpus. We followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) [193] to identify and select the relevant papers (Fig. 1). The original search yielded a corpus of 266 unique papers,

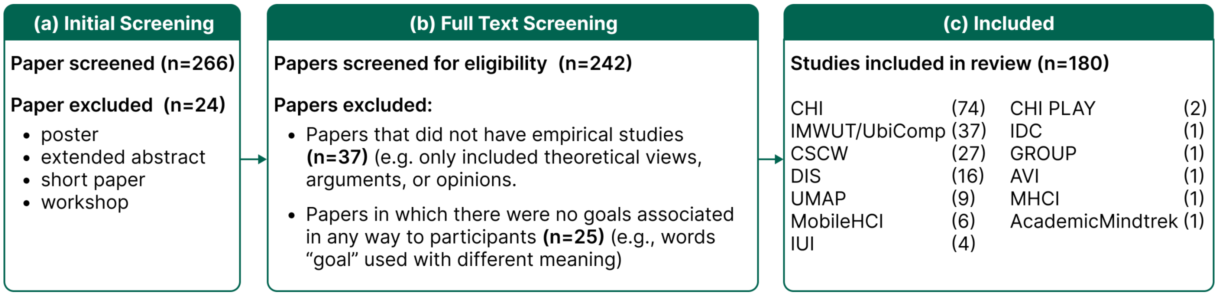

Figure 1: Selection criteria of the paper corpus following the PRISMA guideline.

which appeared in 13 different venues. In the initial screening (a), we excluded papers that were non archival (N=24), such as posters or workshops. From the 242 papers that we fully screened, (b) we further excluded papers not describing empirical studies, such as those solely focused on theory, arguments, or opinions (N=37), and those in which references to goals did not have the meaning of a goal pursued by participants (e.g., where the words related to "goals" might refer to research team goals) (N=25). This led to (c) 180 papers, which were included in our final review (indicated by * in References, although not all of them are cited in our paper).

The first author downloaded each paper from the initial set (N=266) and completed the screening against the identified criteria, a process that took place over a few months with weekly input from two senior researchers.

3.2 Coding Process

The process of coding these 180 papers took place over two years and involved five coders and two senior researchers across two continents, who met weekly or more to discuss and support the coding process and used a dedicated Slack channel to discuss and reach an agreement (Fig. 2). The coding process was hybrid [72], involving both deductive and inductive codes relevant to our research questions, and consisted of two stages: (i) codebook development and (ii) codebook use and revision, throughout which we developed and tested both the initial and the revised codebook.

First stage: Codebook development (Fig. 2, Stage 1) involved identifying the deductive codes informed by concepts and theories from prior research such as types of research contributions [259], or types of interventions for behavior change [170]. With the initial codebook of deductive codes (Codebook V1), we coded half of the corpus while also generating the inductive codes as we drew insights from the papers, such as the different types of stakeholders. Inductive codes had a broad set of values reflecting text segments from papers, which we worked on organizing as closed codes. At the end of this stage we generated the codebook with both deductive and inductive codes (Codebook V2), whose quality we tested on 25 papers randomly selected from the remaining half of the corpus. The codebook testing involved four pairs of independent coders coding between 5 and 10 papers, in total N=25, an adequate sample size for testing codes with many values like ours [28]. Inter-rater agreement was computed using Cohen's Kappa [43] and Monte Carlo simulations to compute the chance agreement components of kappa on codes with multiple item selections, following previous similar approaches [230]. Findings showed 0.51 Cohen's Kappa with a substantial agreement for 47% of codes, moderate/fair for 37% of codes, and 16% of codes showing low or no agreement. All codes with a moderate agreement or less were revised or discarded. We also had each of these revised codes allocated to individual coders who rechecked them within the set of papers already coded. The codebook addressed RQ1-RQ3.

Second stage: Codebook use and revision (Fig. 2, Stage 2). With the revised codebook, we coded all, but 25 papers from the remaining half of the corpus. The 25 papers were used to compute the second inter-rater agreement. Findings showed 0.72 Cohen's Kappa with substantial agreement on 76% of codes, moderate and fair on 21% of codes, and slight agreement on 3 of the codes that were revised. This led us to a total of 50 papers, representing over 25% of the corpus, on which we ensured the overall evaluation of the codebook.

In the second stage, we followed the PRISMA guidelines and conducted an additional screen to include papers published between August 2022 and June 2023, ensuring the review was updated with the

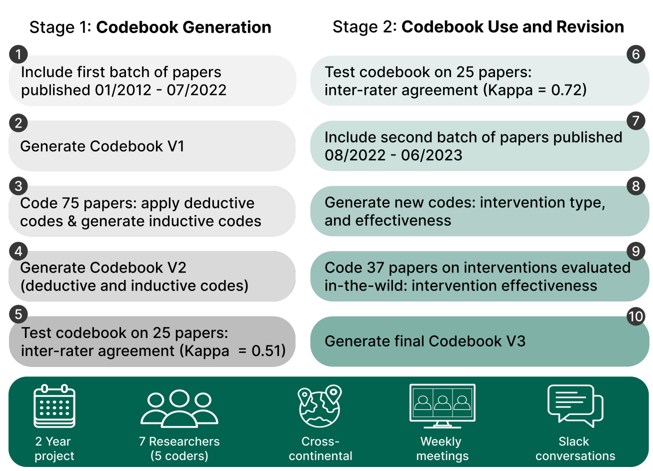

Figure 2: Coding process for systematic review of 180 papers and meta-analysis of 37 papers which explored effectiveness of technology-based interventions in-the-wild. It consisted of two stages: codebook development (left) and codebook use and revision (right), conducted by a team of seven researchers over two years.

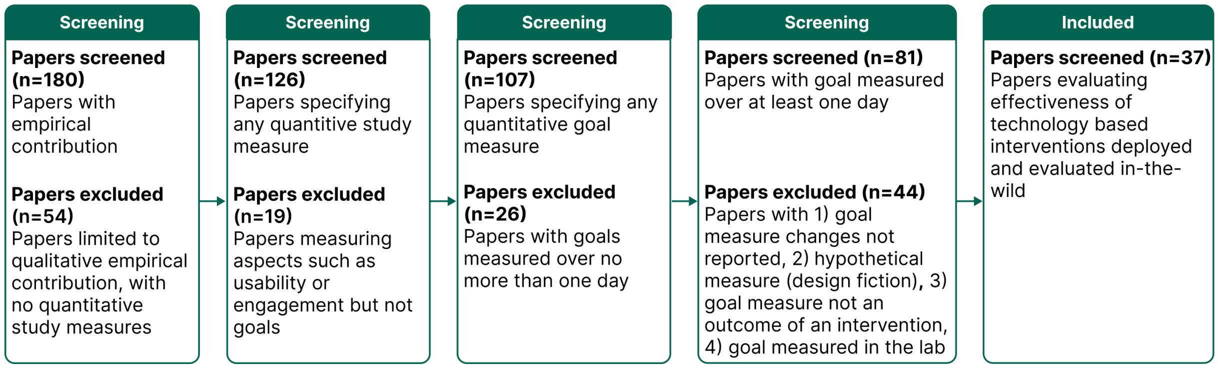

Figure 3: Selection criteria for the 37 papers that reported technology-based interventions deployed and evaluated in-the-wild.

most recent research. We also extended the codes with additional ones relevant to RQ4 on the effectiveness of interventions. Thus, 81 papers were further coded for intervention type as these were described in the lab or in-the-wild evaluation of technology-based interventions (Fig. 3), and for 37 describing technology evaluation in-the-wild, we also extracted data on intervention effectiveness such as study sample size, and change in intervention outcomes. For this, one coder coded all these papers to ensure consistency, while two other coders coded independently for intervention type 10 of these papers (over 10%), reaching agreement through discussion. At the end of this stage, we had the final Codebook V3.

3.2.1 Selection of Subset of Corpus Papers for Meta-Analysis. Given the different types of interventions studied in the corpus and their measures, we wanted to understand which interventions are most effective and for what goal domains of behavior change. To explore this question, we employed a meta-analysis whose aim was to explore the synthesis of the effectiveness across different interventions [92]. From the 180 papers in our corpuswe focused on those deployed and evaluated in-the-wild. The decision relates to our RQ4 on the effectiveness of technology-based interventions. For the definition of effectiveness, we followed Gartlehner et al.'s [80] systematic review where they distinguish *effectiveness* as evaluation in-the-wild, from efficacy or evaluation in the lab.

Therefore, from the 180 papers selected for the systematic review, we followed the following four exclusion criteria: (i) used only qualitative measures (N=54), (ii) used quantitative ones but only for system usability or user engagement rather than for goal measurement (N=19), (iii) those involving goal measures but technologies were used for less than one day (N=26), and (iv) those with goal measures which were not fully reported, were used in design fiction, were not used to measure an intervention outcome or were measured in the lab (N=44). This led to a set of 37 papers (marked by ** in References), which describe technology-based interventions for behavior change, and Figure 3 presents the PRISMA diagram for the selection of these papers.

3.3 Overview of Main Codes

Since our coding process led to a large number of codes, to support the reader, we offer here a succinct overview. Subsequently, in the Findings sections, we provide contextualized, piecemeal introductions of each subset of codes relevant to the specific findings presented in that section, together with those codes' definitions. Thus, Fig. 4 presents an overview of the main codes and extracted data mapped to the research question (RQ1-RQ4) they contribute to.

Thus, RQ1 related codes namely goal domains, and goal characteristics (sociality and motivation) are described in 4.1.1 and 4.1.2, respectively. The RQ2 related codes include HCI main contributions in 4.2.1, research methods in 4.2.2, research approach and measures in 4.2.3, and stakeholders in 4.2.4. For RQ3, the main codes include technologies 4.3.1, underpinning theories 4.3.2, as well as evaluation site and duration in 4.3.3. Finally, for RQ4, the main codes are intervention types detailed in 4.4.1, while extracted data is described in 4.4.2. Note that papers reporting multiple studies were coded separately for each study.

4 FINDINGS

The findings are organized under four sections matching research questions (Fig. 4): goal domains and characteristics; goals studied capturing HCI contributions, research methods, approach, measures, and stakeholders; studied technologies including their theoretical underpinnings and evaluation; and the effectiveness of technology-based interventions in-the-wild.

4.1 User Personal Goals: Domains and Characteristics

This section describes the user's personal goals, their domains and characteristics, such as sociality and motivation (RQ1).

4.1.1 Goal Domains. Goal domains captured the domain of the goal that the paper discusses. Inductive coding led to domains such as physical health and wellbeing (e.g., physical activity, nutrition), mental health and wellbeing, digital wellbeing, productivity, learning, sustainability, or finance. In alignment with previous work [62, 68], our findings emphasize health and wellbeing as main domains, providing a more nuanced understanding of goal-focused research in behavior change.

First, within the prevalent health and wellbeing domain (80.4%), findings indicate two unequally represented areas, with physical health and wellness being over three times larger (66%) than mental

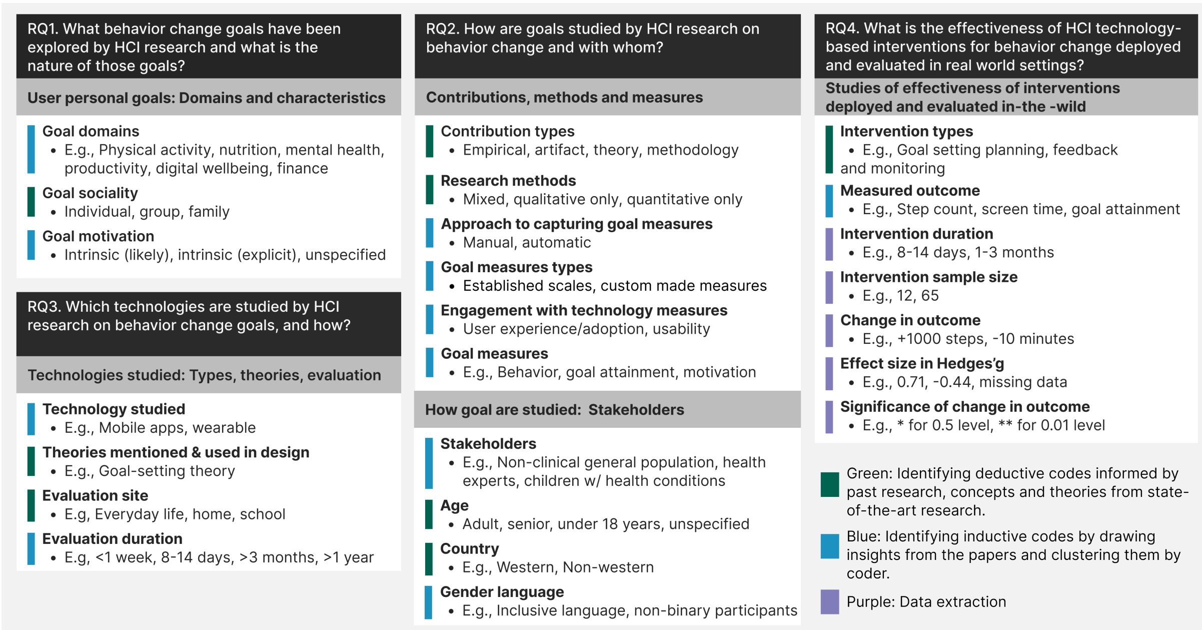

Figure 4: Overview of generated main codes and extracted data, mapped to research questions. The color-coded vertical bars indicate the process of code generation (green: deductively generated, blue: inductively generated, purple: extracted data.)

health and wellbeing (14.5%). Figure 5 shows the distribution of papers within single or multiple domains, with each paper being counted once within the chart (total 100%). Table 1 presents a complementary take, as frequencies of goal domain codes to account for all domains targeted by each individual paper.

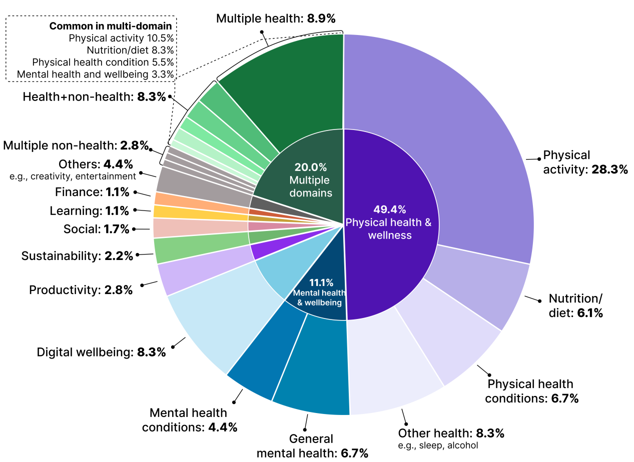

Figure 5: Distribution of papers based on goal domain with a focus on single vs multiple goal domains represented. Most papers target goals within a single domain (80%) such as physical activity, or nutrition, or physical health, while the rest (20%) focus on multiple goal domains where health is also predominant. Each paper is counted only once.

Second, within the overall health and wellbeing domain, research interest has been by far on *physical activity* (N=70, 38.9%), followed by nutrition and diet (N=26, 14.4%), and physically diagnosed conditions (N=22, 12.2%). Within physical activity (N=70), 28.3% of papers focused on physical activity only, and 10.5% on multiple domains. The most common physical activity goals include exercising, usually measured as step counts (N=29) [13, 30, 172, 174, 178, 208], which illustrate *quantitative goals* defined as goals that can be directly tracked [185]. Papers related to *nutrition and diet domain* (N=26, 14.4%) include a subset of 6% that focused exclusively on diet, with goals targeting healthy eating [139], weight [98], or diet management [153, 224]. The domain of *physically diagnosed conditions* (12.2%) (7% single domain, 5% in multi-domain papers) targeted mostly diabetes [59, 105], asthma [175], and chronic condition management [61]. *Other physical health and wellness* papers (N=23, 12.8%) focus on alcohol intake [191], sleep [104, 134], or rehabilitation [251].

Within the *mental health and wellbeing* main domain (N=26, 14.4%) (11% focus only on mental health, 3.3% on multiple domains), findings indicate two different areas. One focuses on *general mental wellbeing* (N=16, 8.9%) goals related to reducing stress [200], or avoiding procrastination [35], while the other targets goals for *mental health diagnosed conditions* (N=10, 5.6%) for managing attention deficit hyperactivity disorder [41], anxiety disorder [229], or diagnosed depression [71]. Another important outcome is that diagnosed conditions for both physical (N=22, 12.2%) and mental health (N=10, 5.6%) have been targeted the least within the health and wellbeing domains. This is surprising, given the growth of HCI in affective health [220], indicating the value of stronger interaction with behavior change research.

Table 1: Distribution of papers with respect to goal domains. Some papers are coded with multiple subcodes within a main code.

Code	Frequency		Code	Frequency	
Goal Domains					
Physical activity	70	38.9%	Mental health diagnosed condition	10	5.6%
Nutrition/diet	26	14.4%	Social interaction	7	3.9%
Other physical health	23	12.8%	Sustainability	7	3.9%
Physical diagnosed condition	22	12.2%	Learning	7	3.9%
Digital wellbeing	19	10.6%	Finance	5	2.8%
General mental wellbeing	16	8.9%	Others	7	3.9%
Productivity	13	7.2%			

Our outcomes also highlight a growing interest in other domains, most notably *digital wellbeing* (N=19, 10.6%), followed by goals targeting *productivity* (N=13, 7.2%), *sustainability* (N=7, 3.9%), *social interaction* (N=7, 3.9%), *learning* (N=7, 3.9%), and *financial* (N=5, 2.8%) domains. This is an important finding extending the focus on specific goal domains, less explored in previous reviews [62, 68]. Digital wellbeing targeted limiting phone use [194], or reducing time spent on social media [123], representing also growing topic in other HCI areas such as technologies for self-regulation [11, 158]. The productivity domain has looked at time management [156], or increasing productivity while multitasking [132].

Findings indicate limited research on behavior change in other domains, with an emerging interest in *sustainability* (N=7, 3.9%), *social interaction* (N=7, 3.9%), *learning* (N=7, 3.9%), or *finance* (N=5, 2.8%) domains. Examples of studies that focused on such goals include reducing electricity consumption [69, 115], saving energy during laundry [90], families sharing food-related photos for social support [154], increasing social connectedness [58], or writing skills to foster self-regulated learning [256], and discouraging impulsive online spending [177]. The sustainability and financial goals reflect more recent yet important HCI areas, so their highlight by our findings merit acknowledgment.

Single vs multiple domain goals. We also looked at if and how the reviewed papers explore multiple goals, within or across domains. Findings indicate that only 36 papers (20%) target goals from multiple domains, usually two (N=26) and less commonly three (N=6), four (N=2), five (N=1), or six domains (N=1). The most prevalent co-occurring domains are physical activity (N=19, 10.5%), diet (N=15, 8.3%), physical health conditions (N=10, 5.5%), and mental health/wellness (N=6, 3.3%). Almost half of these papers (N=16) include only goals across health-related domains. Most of the other half (N=15) include goals from at least one of the health-related domains and one from non-health domains, such as sustainability and nutrition/diet (N=1), physical activity and productivity (N=1), finance, physical activity, and nutrition/diet (N=1). The remaining papers (N=5) include goals across non-health domains such as digital wellbeing and learning (N=2), and productivity and learning (N=1). Another interesting outcome is that while some papers (N=7) strive to integrate goals such as productivity and wellbeing at work [99], or physical activity to manage chronic conditions [131], others target goals from multiple domains in isolation, i.e., saving money and exercising [6].

Key takeaways: Strong emphasis on physical health and wellness domains, particularly physical activity, followed by mental health and wellbeing domains, including diagnosed physical or mental conditions. There is limited focus on other domains besides health and wellbeing, mostly digital wellbeing. Most papers target goals within single domains, with only 20% exploring goals across domains, usually 2, less frequently 3-6 domains, with at least one of them being health or wellbeing. Few papers target the integration of such goals rather than merely exploring them in isolation.

4.1.2 Goal Characteristics: Sociality and Motivation.

Goal sociality. To code for goal sociality, we employed Chulef et al.'s distinction [38] between individual (or intrapersonal), and social (or interpersonal) goals related to family, or broader social goals. Findings confirm the prevalence of individual goals (N=151, 83.9%) [68], but also the small, albeit growing interest in social goals (N=34, 18.9%) (Table 2). Among the latter, two-thirds are goals related to social groups like communities (12.2%), and the remaining third to families (6.7%). Findings also indicate that only a few papers focus on collective goals or one unique goal pursued collaboratively, while most papers in the social domain are merely shared or identical goals pursued independently by each individual [233]. For example, shared family goals include awareness of and reduction in household energy consumption [115], which however, may not be a group goal since the recruitment criteria required that at least one, but not each household member uses the app. An example of a group goal is working together towards a community display aimed to promote physical activity in local neighborhoods

Table 2: Distribution of papers with respect to goal characteristics: sociality and intrinsic motivation (N=180). Some papers fit multiple subcodes within a main code.

Code	Frequency	
Goal Sociality		
Individual	151	83.9%
Group	22	12.2%
Family	12	6.7%
Goal Motivation		
Intrinsic (likely)	63	35.0%
Intrinsic (explicit)	41	22.8%
Unspecified	76	42.2%

Table 3: Distribution of papers with respect to HCI contributions (N=180). Some papers fit multiple subcodes within a main code.

Code	Frequency	
Contribution Types		
Empirical	180	100.0%
Artifact	110	61.1%
Theoretical	9	5.0%
Methodological	8	4.4%

[84]. We also looked at how individual/social goals vary by domain, and findings indicate that while individual goals are prevalent in most domains (over 80%), social ones are more common in the sustainability (5/7) domain. This outcome highlights the value of studying collective goals, such as generative concerns for younger generations which underpin environmental sustainability [10].

Goal motivation. We coded for goal motivation if recruitment criteria explicitly required participants' interest in the goal targeted by the technology being explored, i.e., *intrinsic explicit*, or if recruitment criteria required participants whose goals were likely aligned to those targeted by the explored technologies, albeit not explicitly required, i.e., *intrinsic likely.* An important outcome is that the largest set of papers (N=76, 42.2%) neither specify nor account for users' intrinsic motivation for the targeted goals (Table 2). Also significant is that less than a quarter of papers (N=41, 22.8%) explicitly recruited participants whose personal goals matched those targeted by the technologies explored in the study. For instance, Konrad et al. [120] recruited participants interested in reducing their stress. Other efforts to match the goals targeted by technology are reflected in the second largest group of papers (N=63, 35%), which use recruitment criteria as a proxy for participants' potential intrinsic motivation in such goals. Among these, some papers (N=23) do not explicitly clarify whether participants were actually interested in pursuing that goal. Proxy criteria included, for instance, health conditions that made it likely for people to be motivated to pursue the goal studied by researchers, such as Type 2 Diabetes for studies related to nutrition goals [78], participants living with depression for studies on mental health apps [121], or those managing multiple chronic conditions for studies on self-management technologies [61]. Apart from intrinsic motivation for goals targeted by technology, some papers also explored users' motivation for behavior change by employing the Transtheoretical Model (TTM) [206] for screening off participants not ready to engage in the goal being studied [194].

Key takeaways: Strong emphasis on individual goals, with limited focus on social goals. The latter tend to be shared rather than collective or collaboratively pursued. Limited explicit focus on users' intrinsic motivation for goals, which is tacitly assumed through recruitment.

4.2 Goal Studied: Research Contributions, Research Methods, Measures and Stakeholders

This section describes how goals are studied with a focus on main HCI contributions, research methods, measures, and with whom, i.e., stakeholders (RQ2).

4.2.1 Contributions to HCI Research. Using Wobbrock and Kientz' classification [259], findings indicate that all papers (100%) make empirical contributions, followed by artifacts (61.1%), and to a lesser extent theoretical (5.0%) or methodological (4.4%) contributions (Table 3). The prevalence of *empirical contributions* is not surprising, given our inclusion criteria for studies of participant goals. A third of the papers made only an empirical contribution (N=59, 32.1%), while the rest of the papers (N=121, 67.2%) made *multiple types of contributions*, most often reporting both empirical and artifact contributions (N=100, 55.6%). As we excluded papers that did not involve participants, our corpus did not include papers that only make an opinion, literature survey, or theoretical contribution.

Artifacts are the second main contribution reported in over 60% of the papers, which consist of interactive prototypes [64, 95] such as mobile apps (N=59 of 110, 53.6%), web-based applications, or browser extensions (N=18 of 110, 16.4%), wearables and smart watches (N=13 of 110, 11.8%). *Theoretical contributions* featured in 9 papers (5%) and consisted of new models, frameworks, principles, or contributions to design research theories. Examples included models for tracking to support healthy eating [152], for decision-making in self-management of health conditions [105], and for engaging users in mental health therapy [229]. *Methodological contributions* were reported in 8 papers (4.4%) and consisted of novel design methods such as approaches exploring the value of domestic technologies for personal goals [27], design tools for personalizing health wearables [15], or design research methods for tracking practices [85].

Key takeaways: Prevalent empirical contributions, followed by artifact contributions, with considerably less focus on theoretical or methodological contributions.

4.2.2 Research Methods. To code research methods, we employed a classification previously used in HCI systematic reviews [219] consisting of qualitative, quantitative, and mixed methods. Table 4 Research Methods (top left). Most papers employ mixed qualitative and quantitative methods (N=97, 53.9%), with about a quarter (N=47, 26.1%) using only qualitative methods (usually interviews), while another quarter (N=36, 20%) employs only quantitative methods, in particular experiments.

With regard to the qualitative methods, the most common ones are interviews (N=126, 70.0%) to gather participants' needs and experiences both in the early stages of design, as well as their feedback on high-fidelity, interactive technologies [100, 140, 215]. These were followed by participatory design or co-design workshops (N=25, 13.9%) to create prototypes[131], and gather users' feedback [36]. Less reported qualitative methods include focus groups (N=12, 6.7%), diary studies (N=11, 6.1%), or observations (N=7, 3.9%). In terms of quantitative methods, experiments were prevalent (N=48, 26.7%).

Table 4: Distribution of papers with respect to research methods, approach, measure type, goal measures, and engagement measures (N=180). Some papers fit multiple subcodes within a main code.

Code	Frequency		Code	Frequency	
Research Methods			**Goal Measures**		
Mixed	97	53.9%	Behavior	78	43.3%
Qualitative only	47	26.1%	Physical	30	16.7%
Quantitative only	36	20.0%	Digital wellbeing	21	11.7%
Approach to Capturing Goal Measures			Eating	17	9.4%
Manual	79	43.9%	Sleeping	5	2.8%
Automatic	51	28.3%	Restorative	4	2.2%
Not applicable	74	41.1%	Sustainable	1	0.6%
Goal Measures Types			Goal attainment	26	14.4%
Custom made measures	76	42.2%	Motivation	64	35.6%
Established scales	47	26.1%	Emotion	19	10.6%
Not applicable	74	41.1%	Attention	1	0.6%
Engagement with Technology Measures			Body physiology	2	1.1%
User experience/adoption	55	30.6%	Goal-oriented work	9	5.0%
Usability	16	8.9%			
Not applicable	115	63.9%			

Key takeaways: Most papers employed both qualitative and quantitative research methods, most commonly interviews and experiments, respectively.

4.2.3 Goal Measurement: Approach, Measures. To capture the rich range of goal measures, we employed the following codes reflecting the content being measured: behavior, motivation, emotion, attention, body-related metrics, as well as goal-oriented work. We also coded the *approach used to capture goal measures* if such measures were captured manually by participants or automatically logged (e.g., step counter, use logs). *Goal measure types* were coded as such if the reported measure was established, such as a validated scale or a custom-made one.

Findings indicate that most studies captured goal measures manually through self-reports (N=79), with under one-third using automated means (N=51). We also found wide use of established scales (N=47), most validated in prior research (N=43). Supplementary material, Section 5 lists all validated scales for goal measures across domains and curated references reporting their validity.

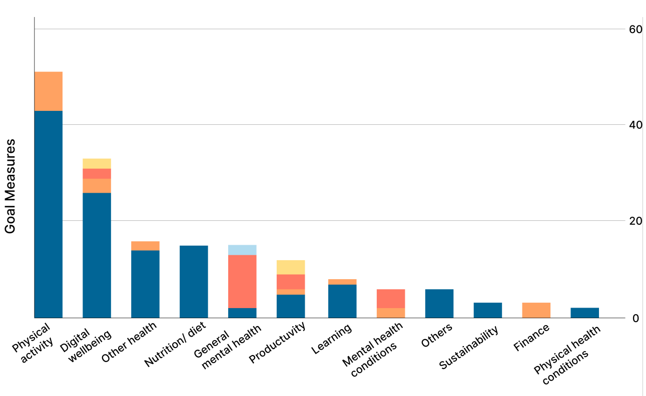

Figure 6: Distribution of papers with respect to goal measures broken down by goal domains.

Regarding measured content, findings indicate a broad range of goal measures. Figure 6 shows the use of various goal measures across domains, with behavior measures being prevalent (N=78) across most domains, emotion measures most commonly used in mental health general and diagnosed condition domains, and attention mostly in productivity and digital wellbeing domains.

We organized *behavioral measures* into eight groups related to (i) physical activity (i.e., step count) (N=21), duration (N=3), distance (N=2), (ii) phone overuse (i.e., screen time) (N=12), count of phone/app use (N=5), problematic behavior (N=2), (iii) eating behavior (i.e., food consumption) (N=5), food variety (N=3), calorie count (N=2), or portion size (N=1), (iv) sleep behavior such as duration (N=3) and quality (N=2), (v) learning behavior (N=6), (vi) restorative behaviors such as duration spent in nature (N=4), and (vii) sustainable behavior such as energy consumption (N=2). The eighth behavioral measure relates to the goal itself, namely goal attainment. *Goal attainment* measures, such as simple, self-reported measures of goal progress or completion, are a surprisingly popular behavior-agnostic measure across domains (N=30). Goal attainment scales were often reported in categorical format, i.e., yes or no options to answer if the goal has been completed [120, 138], followed by self-reported percentage of goal completion, i.e., 50% [101, 208], and 9- or 10-point Likert scales [82, 134].

Motivation-related measures are the second largest after behavior measures, including (i) motivational beliefs such as self-reported self-efficacy [22, 203] (N=6), (ii) motivation for change informed by the Transtheoretical model [206] (N=6), and (iii) other aspects such as self-control, perceived fit for self, reward, coercion, or persuasion (N=13).

Emotion-related measures include self-reported negative emotions such as stress or anxiety, most often through validated scales such as the Patient Health Questionnaire (PHQ-9) or Perceived

Table 5: Distribution of papers with respect to stakeholders, age, country and gender language (N=180). Some papers fit multiple subcodes within a main code.

Code	Frequency		Code	Frequency	
Stakeholders			**Age**		
Non-clinical general population	111	61.7%	Adult	157	87.2%
Adults w/ health conditions	29	16.1%	Senior	52	28.9%
University students	24	13.3%	Under 18 years	18	10.0%
Health experts	16	8.9%	Unspecified	29	16.1%
Education experts	9	5.0%	**Country**		
Workers	7	3.9%	Western	87	48.3%
Children w/ health conditions	4	2.2%	Non-western	19	10.6%
Students - pre-college	4	2.2%	Western & Non-western	4	2.2%
Design experts	3	1.7%	Unspecified	70	38.9%
Unspecified	4	2.2%	**Gender Language**		
			Non-inclusive language	141	78.3%
			Inclusive language	23	12.8%
			Non-binary participants	21	11.7%

Stress Scale (N=15), and positive ones such as wellbeing or life satisfaction (N=4).

Attention measures include self-reported mindfulness, absent-mindedness, or focused attention (N=4), multitasking (N=1), and task time (N=1). In contrast, *body physiology related measures* such as heart rate or physical symptoms, were surprisingly few (N=2), and employed in health and wellbeing domains.

Goal-oriented work. We also noted the emergence of goal-oriented work [27], for instance, the use of Personal Project Analysis (PPA) workbooks [146, 147]; with PPA being a cognitive motivational method commonly used in both therapy [55], and health domains [48] to support people accessing, and reflecting on their goals. Goal-oriented work also included breaking down the goals and prompting readiness to change in therapy domain [35] facilitated by motivational interview [212]. Another example is Chaudhry et al.'s [33] use of a goal-oriented care approach through which health experts used motivational interviewing to enable patients living with multiple conditions to identify the goals they would like to pursue, followed by setting SMART goals and tracking them. Lee et al. [137] employed a similar approach by exploring health experts' work of facilitating users' deeper reflection on their own goals. Finally, under goal-oriented work, we also have scales for self-reported goal criteria such as importance and difficulty (N=3).

Engagement with technology measures. An interesting outcome is the use of behavior measures as indicators for engagement with the technology (N=55, 30.6%), including long-term adoption in the wild. For instance, through measures of physical activity, as well as system logs on technology use. In addition to goal measures, other quantitative measures include traditional usability scales (N=16, 8.9%) such as SUS, SASSI, and NASA TLX.

Key takeaways: Goal measures are captured mostly through self-reports such as validated scales and custom-made measures. Measures target predominantly behavior for tracking physical activity, screen time, food consumption, sleep, learning, restorative, or sustainable behaviors, followed by goal attainment, motivation, emotion, and to a lesser extent, attention and body physiology measures. Nascent research applies goal-oriented methods to support awareness and reflection on goals and their deep motives.

4.2.4 Stakeholders. We inductively generated codes from recruitment criteria such as diagnosed conditions, participants' occupation, and age (using the World Health Organization guidelines [260]). Findings indicate that most papers (61.7%) engaged with non-clinical general population (N=111), followed by adults diagnosed with health conditions (N=29), and university students (N=24) (Table 5). A significant outcome is a limited engagement of health experts such as medical staff, therapists, eating or sleep experts, with only 16 papers engaging them, despite the largest interest in physical or mental health/wellbeing. Findings also indicate a small set of papers involving children with diagnosed physical (N=2) and mental (N=2) conditions. Additional groups of participants included education experts (N=9), workers (N=7), school children (N=4), and design experts (N=3).

Vulnerable participants: Clinically diagnosed people, children, or older adults. We also looked at the inclusion in participant samples of vulnerable users namely clinical populations, children or older adults. Findings highlight that from papers that engaged with participants living with diagnosed conditions, the most commonly reported were diabetes (N=4) [78, 105], asthma (N=2) [175, 252], multiple sclerosis (N=2) [20, 82], migraine (N=2) [227, 228] and other conditions (N=7) such as cancer and heart disease [164, 266]. In addition, 6 papers targeted the management of multiple chronic conditions [33, 61, 138]. In the mental health domain, papers reported studies with participants living with diagnosed depression (N=4) [71, 169], depression and anxiety (N=3) [121, 264], ADHD (N=2) [41, 239], or psychosis (N=1) [70]. With respect to age, 157 papers (87.2%) included adults between 18–60 years old. Two sets of papers engaged vulnerable users, namely children or teenagers under 18 years of age (N=18), and senior adults over 60 years (N=52) including those with diagnosed conditions (N=16).

Participant samples: Inclusiveness. Inclusiveness of participant samples matters as it reflects *with whom* and *for whom* the HCI

Table 6: Distribution of papers with respect to types of technologies explored and their theoretical underpinning (N=180). Some papers fit multiple subcodes within a main code.

Code	Frequency		Code	Frequency			
Technology Studied			**Theories**	**Mentioned**		**Used**	
Mobile apps	78	43.3%	Goal-Setting Theory	32	17.8%	12	6.7%
Wearable	29	16.1%	Transtheoretical Model	22	12.2%	8	4.4%
Web/Tablet	24	13.3%	Social Cognitive Theory	17	9.4%	6	3.3%
Tangible	9	5.0%	Self-Determination Theory	16	8.9%	10	5.6%
Ambient	8	4.4%	Five-Stage Model	15	8.3%	2	1.1%
Unspecified	32	17.8%	Lived Informatics Model	15	8.3%	3	1.7%
			Dual Systems Theory	10	5.6%	5	2.8%
			Theory of Planned Behavior	8	4.4%	N/A	N/A
			Fogg Behavior Model	8	4.4%	3	1.7%
			Social Comparison Theory	5	2.8%	4	2.2%
			Self-Regulation Theory	5	2.8%	3	1.7%
			Others	121	67.2%	51	28.3%
			Unspecified	47	26.1%	103	57.2%

research on goals for behavior change is undertaken. To understand if participant samples were broadly inclusive and representative, or if they reflect the acknowledged HCI WEIRD bias (Western, Educated, Industrialized, Rich, Democratic) [144], we employed Linxen et al.'s [145] categories of Western and non-Western focus of the research to code for the country in which the study took place. We also coded for the gender language of reporting participant samples such as sex, gender, or more inclusive language. Our outcomes confirm the WEIRD bias, as most papers (61.1%) among those which reported participants' country (N=110), mentioned the Western context (N=87) in which they were conducted, namely the United States (N=61), followed by the United Kingdom (N=14), Germany (N=5), and Canada (N=5). The remaining 19 papers reported non-Western countries namely Korea (N=8), China (N=4), and India (N=3). With respect to gender-inclusive language, only 21 (11.7%) papers included participants who identified as non-binary (e.g., trans, genderqueer). Most papers reported binary gender categories: 23 papers reported women/men, while the largest number of papers (141, 78.3%) reported biological sex (female/male).

Key takeaways: Prevalent engagement with nonclinical population, albeit limited with the clinical population (i.e., diabetes, depression, anxiety) or experts. Vulnerable users included mostly older adults, followed by children or teenagers. Limited inclusiveness of participant samples, confirming WEIRD and binary gender biases.

4.3 Technologies Studied in Research on Goals for Behavior Change: Types, Theories, Evaluation

This section describes technologies studied by HCI research on goals for behavior change, their theoretical underpinning, and the site and duration of their evaluation.

4.3.1 Types of Technologies. We coded the types of studied technologies as mobile app, wearable, web application, browser extension, tangible, ambient device, tablet/desktop. Findings show that most papers (N=148, 82.2%) explore specific technologies, while others (N=32, 17.8%) focus on broader classes of technologies such as designing for mental wellness [106], or models such as that of personal informatics [67].

Table 6 (left) shows types of technologies and their frequencies. *Mobile apps* are by far the most explored technologies (N=78, 43.3%) across domains, followed by wearable (N=29, 16.1%), web/tablet apps (N=24, 13.3%), and to a lesser extent tangible (N=9, 5.0%), and ambient technologies (N=8, 4.4%). Mobile apps focusing on health and wellbeing include those motivating daily exercise [101], management of stress [245], or chronic conditions [173]. Examples from other domains include apps supporting work productivity [118], or energy saving [140]. *Wearables* such as smartwatches and trackers are mostly used for health and fitness [198], or wellness like stress tracking [183]. Interestingly, *web-based applications and browser extensions* feature almost as much as wearables. However, unlike mobile apps prioritizing fitness goals, web-based apps and browser extensions are preferred for wellbeing or productivity goals [99, 123]. While less explored, *tangibles* focused mostly on health/wellness, sustainability [223], or learning goals [100]. *Ambient technologies* include smart home and IoT devices targeting mostly sustainability, learning, and digital wellbeing goals.

Key takeaways: Most papers focus on specific technologies in particular mobile apps and wearables for physical activity, health, and wellbeing goals, and less so on tangible or ambient technologies which target sustainability, learning, and digital welbeing goals.

4.3.2 Theoretical Underpinning. We coded whether papers engaged with theory in related work, i.e., theory mentioned, or in the design of artifacts, i.e., theory used. An important outcome is the limited use of theories for design (N=77, 42.8%), despite extensive references to them in papers' state-of-the-art (N=133, 73.9%), and the large number of different theories both mentioned (121 theories) and used (57 theories), with over 70% of them being used only once. The top four most used theories are described below.

Goal setting theory [148] argues for the motivational value of setting, tracking, and monitoring appropriate goals, which should

Table 7: Distribution of papers with respect to technology evaluation site, and duration (N=125). Some papers fit multiple subcodes within a main code.

Code	Frequency		Code	Frequency	
Evaluation Site			**Evaluation Duration**		
Controlled setting	24	19.2%	<1 day	13	10.4%
Everyday life	70	38.9%	2-7 days	9	7.2%
Home	12	6.7%	8-14 days	23	18.4%
School	6	3.3%	15-21 days	18	14.4%
Workplace	5	2.8%	22-31 days	20	16%
Church	2	1.1%	1-3 months	23	18.4%
Hospital	2	1.1%	>3 months	9	7.2%
Others	4	2.2%	>1 year	1	0.8%
Unspecified	3	1.7%	Unspecified	14	11.2%

be specific and somehow challenging as they lead to higher performance, which in turn supports self-efficacy. Among our reviewed papers, it was used for physical activity, digital wellbeing, or nutrition goals [34]. *Self-determination theory* argues that by meeting psychological needs of competence, autonomy, or relatedness, people can internalize over time external motivation [57]. It was used for physical activity and social goals: Storywell app encourages parents and children to set physical activity goals, whose completion is rewarded with stories [216]. *Transtheoretical model of change* [206] identifies 5 readiness to change stages: precontemplation, contemplation, preparation, action, and maintenance. Various scales operationalized it to specific domains [163], and reviewed papers use them to screen out participants in the precontemplation stage who are not interested in behavior change [112, 194, 208]. *Social cognitive theory* [21] argues that high self-efficacy, or motivational belief in the ability to complete a goal [203], is a strong determinant of behavior change, which can be supported by experiencing success towards goals. This theory informed Lee et al.' [138] design of a smart pillbox providing visual, near real-time feedback on medication taken to support self-efficacy and self-awareness.

Key takeaways: While most papers mentioned theories, less than half use them to inform technology design, albeit such use is thin, engaging many theories only once, with a small set of theories being consistently used by less than 20% of papers.

4.3.3 Technology Evaluation: Site and Duration. We coded evaluation sites such as labs (controlled environment) or real-world settings (in the wild), including specific sites such as everyday life, schools, and hospitals. Findings show that from the 148 papers studying technologies, 125 reported their evaluation.

Evaluation Site. An important finding is the extensive body of work (N=99, 55%) focused on evaluation in-the-wild, with the remaining papers (N=24, 19.2%) describing *evaluation in controlled settings* such as in the lab (Table 7). Unlike the former group focusing on fully working prototypes, the latter evaluates early-stage prototypes [184] targeting mostly physical activity [64] and digital wellbeing [231], or fully working prototypes whose "preuse acceptability" [181] was explored in the lab, as a prerequisite for later deployment. Findings indicate the prevalence of deployment for *everyday life* (N=70) with technologies targeting goals related to physical activity, nutrition/diet, diagnosed physical and mental health conditions, mental wellbeing, digital wellbeing, or productivity [82, 224]. In addition, papers (N=27) report specific deployment sites namely home (N=12) [69, 138, 261], school (N=6) [7, 160, 256], workplace (N=5) [32, 168, 208], church (N=2) [109, 188], and hospital (N=2) [113, 252], most often targeting domains like health and wellness or diagnosed physical condition for homes, learning for schools, physical activity and nutrition/ diet for workplaces, wellness and wellbeing for churches, and diagnosed physical condition for hospitals.

Evaluation Duration. We coded for duration: under one day, 1, 2, 3 weeks, 1-3 months, over 3 months, or over 1 year, for which we draw on the recommended several weeks duration for evaluation of behavior change technologies [93, 116]. Outcome show that most of the technology evaluation in controlled settings took place within one day (N=13, 10.4%), while deployment in the real-world settings focused on continuous use over longer durations. The latter lasted most often less than one month (N=70 of 99, 70%), of which 61 were evaluated for more than a week, for example, 1 and 2 weeks (N=23, 18.4%) or between 3 and 4 weeks (N= 20, 16%). The next most common duration was between 1 and 3 months (N=23, 18.4%) of 125 real-world deployments. Findings also show that deployments less than 1 week (N=9, 7.2%) or over 3 months are rare (N=9, 7.2%), and those over 1 year are even more so (N=1, 0.8%). In terms of goal domains, deployments between 1 and 3 months targeted mostly physical activity [160, 216], digital wellbeing [123], and sustainability [115]; those between 1 and 2 weeks focused on nutrition/diet [154, 224], other health and wellness [100], and social domain [166], while those between 3 and 4 weeks on physical activity, as well as nutrition/diet.

Key takeaways: Almost 70% of papers evaluate technologies, most of these in-the-wild (N=99), usually lasting between one week and one month, and targeting everyday life or specific sites such as home, school, or workplace, while the rest of papers (N=24) evaluate technologies over less than one day, in the lab.

4.4 Technology-based Interventions for Behavior Change Evaluated in-the-Wild

Our final research question (RQ4) focused on the *effectiveness* of interventions deployed in-the-wild [80]. 81 papers describing evaluations in-the-wild were screened for four criteria described in

Figure 3, leading to 37 papers reporting the evaluation of one or more interventions. For this reason, we pivoted from paper as the unit of analysis, as used in the above sections, to intervention as the unit of analysis, leading to a set of 76 interventions identified within 37 papers.

4.4.1 Intervention Type. By employing Michie et al.'s taxonomy of intervention techniques for behavior change [170], we consistently captured the different types of interventions. Table 8 presents the distribution of codes related to intervention type for the 76 interventions evaluated in-the-wild as identified from the 37 relevant papers. Findings indicate various interventions, most of which, however, belong to a small number of types, as further detailed. *Goals and planning* (N=18/76) and *Feedback and monitoring* (N=13/76) are the two most common interventions across all domains. *Goals and planning* interventions include setting goals for how much time to spend on a digital device [4], or the type of food to eat [224]. Approaches to setting goals include private/public goals [178], or goals adjusted manually/automatically [137]. *Feedback and monitoring* interventions include tracking and reflecting on tracked data, for instance, for wellbeing [97], physical activity [30], or medication intake [138]. *Scheduled consequences* interventions involve withdrawal of something desired albeit problematic, such as access to digital activity. These are the most used interventions for the digital wellbeing domain (N=12/21) to limit screen time [110]. *Shaping knowledge* interventions (N=7/76) provide psychoeducation on wellbeing goals such as sleep hygiene [134], while *Reward and threat* interventions (N=5/76) provide incentives such as financial ones to support sustainability [166], or digital wellbeing goals [194].

Interventions across goal domains. Most interventions targeted physical activity (N=22/76) and digital wellbeing (N=21/76), followed by nutrition/diet (N=8), sleep (N=6), mental health (N=6), sustainability (N=4), and smartphone security (N=4) (see Table 2 in Supplementary material). While the prevalence of some of these reflects the interest of reviewed papers in these domains, the presence of digital wellbeing interventions, those for sleep, sustainability, and cybersecurity are higher than expected. Findings also show different uses of intervention types across domains, with digital wellbeing and nutrition/diet domains benefiting from a larger range of interventions. More importantly, despite various measures of outcome, two domains consistently employ some of these measures, i.e., step count for physical activity, and screen time for digital wellbeing, which allows for a comparative analysis of their effectiveness as shown in our meta-analysis 4.4.3, while other domains do not.

Key takeaways: The most common interventions are goals and planning; feedback and monitoring; and scheduled consequences (N=43/76, 57%). Most interventions target physical activity and digital wellbeing goals (N=43/76, 57%), with consistent use of some measures, i.e., step count or screen time, which allows for comparison of effectiveness, while interventions in other domains employ less consistent measures, which hinder the comparison of their effectiveness. Future work for these other domains can address this gap by focusing on the consistent use of valid outcome measures.

4.4.2 Intervention Effectiveness: Effect Size, Sample Size, Duration, Measured Outcome, Change Significance.

Extracting/computing effect size. We extracted additional data on effect size, i.e., the difference in standard units between the means of the experimental group and a baseline group for all the 76 interventions evaluated in-the-wild. Our analysis indicates that only 14 of these 76 interventions reported effect sizes. In contrast, 30 interventions (N=30/76) reported neither effect sizes nor the data needed to compute them. The remaining 32 interventions did not report effect sizes but provided data on sample size, mean, and either standard deviations or standard errors for both the baseline and experimental phases or groups. Standard errors were converted into standard deviations by multiplying the former by the square root of sample size [222]. For these 32 interventions, we used such data to compute the intervention's effect size in Cohen's d [130], and subsequently Hedges' g as recommended for meta-analysis targeting evaluations of effect sizes for various sample sizes as in our corpus [246]. We used sample sizes to compute the standard error of Hedges' g effect size [246] for the main outcomes targeted by the intervention and assigned statistical significance levels. For the small number of interventions (N=7/76) that have more than one main measure, we also applied the Bonferroni correction [204].

Overview of intervention effectiveness in-the-wild. Figure 1 and 2 in the Supplementary material provide the list of 76 interventions, with additional codes and extracted data as recommended by the meta-analysis [165] namely intervention type and duration, sample size, technology, measured outcome, goal domain, and effect size.

Sample size. Our reviewed interventions were evaluated with participant samples of different sizes (Median=18, Mean=26, Range: 5-169).

Technology. The most common technologies for in-the-wild interventions are still mobile apps (N=38/76) used across domains, followed by web apps (N=13/76) for mental health, physical activity, and sustainability goals; wearables (N=10/76) for physical activity, health, and sleep; and browser extensions (N=7/76) for digital wellbeing.

Intervention duration. 59% of the interventions (N=44/76) were evaluated between one and four weeks, 30% (N=23/76) between 1 and 3 months, while 9.2% (N=7/76) studies deployed for over 3 months, and 2 did not control for the duration.

Significant intervention impact. Among interventions for which we extracted or computed effect size (N=46), 41% (N=19) show significant desired changes in measured outcomes, while the rest

Table 8: Distribution of behavior change technology-based interventions evaluated in-the-wild (N=76) in terms of intervention type

Intervention	Frequency	
Goals and planning	18	23.7%
Feedback and monitoring	13	17.1%
Scheduled consequences	12	15.8%
Other	8	10.5%
Shaping knowledge	7	9.2%
None (control)	7	9.2%
Reward and threat	5	6.6%
Social support	2	2.6%
Associations + rewards	2	2.6%
Associations	2	2.6%

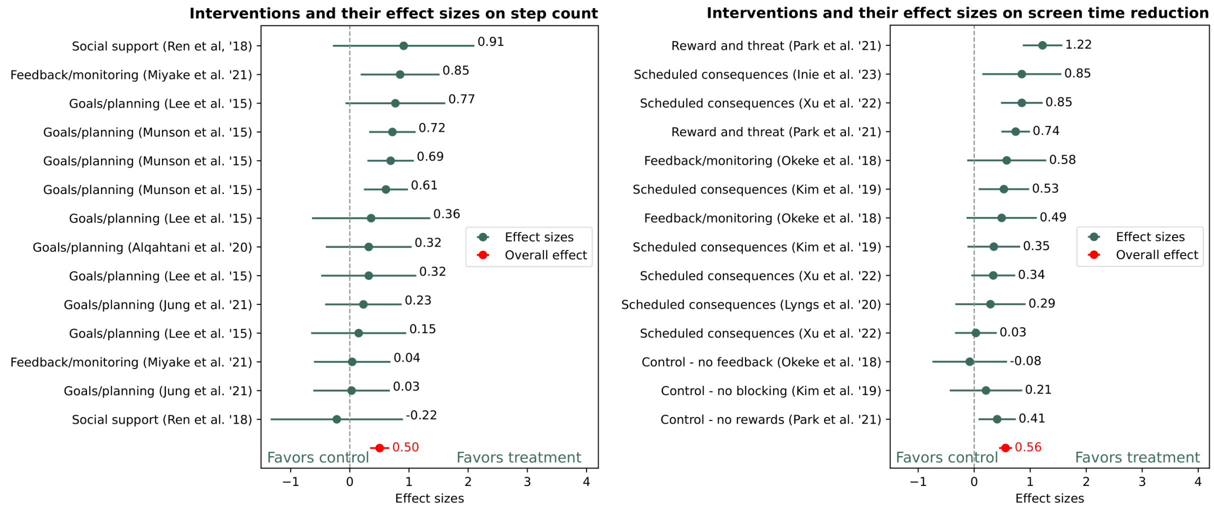

Figure 7: Forest plots showing the interventions and their effect sizes in Hedges' g (mean and 95% confidence interval) sorted in descending order, followed by controls and overall effect (left: interventions on step count, right: interventions on reducing screen time)

either fail to demonstrate significant effects (N=24, 52%) or show significant change albeit in the opposite rather than the desired direction (N= 3, 7%). We detailed five highly effective interventions across domains in the Supplementary material. The lack of significance was associated with low sample sizes [208], different types of interventions where the targeted outcome is not high enough, i.e., moderated goals for physical activities [13], or too high [174], or when the targeted measure of outcome is unfamiliar to users, such as self-reported sleep quality which people develop awareness for and learn to report more accurately,while earlier self-reports are less accurate [50].

Key takeaways: The average intervention evaluated in-the-wild is delivered on mobile app, to samples no larger than 30, using it between 1 and 4 weeks. While 41% of the interventions across domains led to significant desired changes in their measured outcomes, the lack of significance shown by 52% of the interventions suggests the importance of stronger design rationale better addressing, for instance, the issue of goal difficulty and users' familiarity with measured outcomes. At least 39% of the evaluated interventions do not report data on effect size, which hinders the ability to estimate their impact and to contribute to cumulative science [130].

4.4.3 Meta-Analysis: Interventions' Effectiveness. The most common measures of intervention outcomes were screen time (N=19/21) for digital wellbeing, step count (N=18/22) for physical activity, followed by goal attainment measure (N=13/76) across multiple domains, i.e., self-reported goal progress, behavior frequency, or "successful days" of engaging in target behavior [95]. Given the different types of interventions and their measures, it is not trivial to see which interventions were most effective. For this, we run a meta-analysis whose aim is to explore the synthesis of the effectiveness across different interventions [92]. From the 76 interventions, we identified those that have the same measured outcome, which led to 2 sets of 14 interventions each. One set has step count as the measured outcome, and the other one has screen time.

We conducted a fixed-effect meta-analysis, on each of these 2 sets of interventions, weighting by inverse variance. Findings indicate that one model exhibits high heterogeneity: I^2 of 0% for step count, and 62% for screen time. We found a significant overall effect of 0.50 (95% confidence interval from 0.34 to 0.67) for the first set of interventions focused on increasing step count, and a similar significant overall effect of 0.56 (95% confidence interval from 0.45 to 0.67) for the second set focused on reducing screen time. Note that positive effects imply beneficial results, as we report the *reduction* in screen time, and not screen time. Figure 7 shows the two Forest plots, left one for interventions for increasing step count, and right one for those for decreasing screen time. The plots show the explored interventions and their respective effect sizes in Hedges' g (mean and 95% confidence interval), sorted in descending order for the experimental groups, followed by controls and overall effect. A recent meta-analysis on digital self-control tools [210] reported a similar overall effect.

The most effective intervention for increasing step count involves social support from proximal peers in a work setting (0.91 effect size) [208], while the alternative version with distal peers was ineffective (-0.22 effect size, worse than baseline). Other effective interventions include providing a predicted trend of hourly step count based on previous step count [174] and answering reflective questions while setting challenging goals [137]. Among the interventions for reducing screen time, the most effective rewards intervention [194] had an effect size of 1.22 for limiting the use of distracting apps. Another digital wellbeing intervention of scheduling consequences or redirection of activity [96] from time-wasting websites to learning platforms reported the second large size effect of 0.85.

Key takeaways: The technology-based interventions evaluated in the wild, targeting physical activity goals and digital wellbeing goals, were effective as a group. Although there was variability, significant overall medium effect sizes of 0.50 and 0.56 were observed, respectively. Social support, particularly from proximal peers, was the most effective intervention for physical activity, while rewards and threats, particularly the loss of virtual rewards, was the most effective for digital wellbeing.

5 IMPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

We now reflect back on our research questions and highlight the main insights from our findings, both for describing the HCI state-of-the-art research on goals for behavior change, and for charting future research directions. The latter are intended to both expand the design space of these technologies beyond the present prevalent research foci, and to deepen the efforts for more effective technology-based interventions.

5.1 Single-Domain Goals vs Multiple-Domain Goals

People have goals that range in focus from social life (family, romance, friends) to health, learning, education, finance, spirituality, career, and more [38]. Unlike people having varied goals, we were surprised to find an overwhelming amount of papers focused on health in general (80%) and on physical activity (38.9%) more specifically. In contrast, mental health and wellbeing are less represented, and so are goals related to managing diagnosed conditions. While physical activity is clearly important, impacting both physical and mental health, we see opportunities for a higher representation of supporting goals in other aspects of people's lives [38]. Emerging work in other domains such as sustainability, or finance is much needed, suggesting a possible shift towards broader goal domains than previously described [62, 68]. These other goal domains could benefit from other relevant HCI research, not focused on behavior change, i.e., sustainability [89, 162], or money and HCI [9, 199].

Findings also show that most goals were explored in single domains, with only 20% of the papers targeting multiple domains, most of these also prioritizing health. We argue for the value of exploring multiple goals, both individual and social, especially across domains [151] so that they better reflect complex goal hierarchies [38]. For instance, the management of co-morbidities is an area that can very much benefit from a multiple-domain goal approach. The exploration of multiple-domain goals can open up exciting new design opportunities for the most holistic, lived experiences. Future work is needed to explore how to best balance their increased benefits with additional demands for competing resources.

5.2 Individual & Low Intrinsic Goals vs Social & High Intrinsic Goals

Our findings confirm the prevalence of individual goals in HCI research for behavior change [68], while social goals are captured by less than 20% of the papers, with only 6% discussing goals as they pertain to families. Since a large sample of our papers focused on health, we were surprised not to see more social engagement with goals, given past calls to action from family informatics literature [201]. We found it encouraging to see some focus of the community level (e.g., neighborhood) particularly focused on sustainability goals, but we hope that more future studies would tackle urgent issues of sustainability [162]. We also encourage focus at the community or neighborhood levels of health-related technologies. Researchers of physical activity have pointed out unique challenges that people living in low socio-economic neighborhoods could encounter in pursuing physical activity [216, 218]. Broadening certain domains of health to the community level could contribute to a better understanding of how to support community health goals. Given the sociality of human experiences, we encourage researchers to pursue more research that involves understanding and designing for people's goals pursuit in social settings.

Another key finding regards goal characteristics and the limited focus on participants' intrinsic motivation for the goals targeted by the explored technology (23%). According to self-determination theory [57], intrinsic motivation is internal, linked to one's interests and values, and thus a strong predictor for long-term engagement in goal pursuit [75]. This limited focus is surprising, given also the commitment required for behavior change, which suggests a possible disconnect of behavior change technologies from users' values and personal goals. This is a significant outcome, and closely linked to how goals are measured. We argue for the value of new research directions focusing on users' high intrinsic goals. This could include supporting users' awareness of them, and efforts to prioritize them. We also need additional approaches to make goal motivation more relevant in the recruitment of users for studies of technologies targeting behavior change.

5.3 Quantitative Goals, Behavioral Measures (Step Count & Screen Time) vs Qualitative Goals, Broader Measures

As shown by our findings, most behavior change technologies focus on quantitative goals, confirming previous emphasis of tracking technologies on quantitative goals, particularly step count [62]. Our outcomes, however, extend these goal measures with a rich set of behavioral measures, including behavioral measures of physical activity (i.e., step count, duration, distance), but also phone overuse (screen time, count of phone/app use, problematic phone overuse), eating behavior (food consumption, variety, calorie count, portion size), sleep behavior (duration, quality), learning behavior, restorative behavior (duration spent in nature or hobbies), sustainable behavior (energy consumption), and goal attainment. The latter is a much used self-report measure of goal progress or completion which can raise validity issues if not clearly defined [232]. These quantitative goal measures across various behaviors offer a useful vocabulary to support the design of goals and behavior change technologies across domains.

In addition to the prevalent behavior measures, other, less used quantitative measures related to motivation, emotion, attention, and body physiology. We argue for the value of supporting also these latter measures.For instance, biosensors are surprisingly less used in this space but extensively explored for affective health [180, 247, 248] and mindfulness technologies [51, 52]. We would like to encourage more research engaging with the less explored qualitative goals, *prior* to focusing on their associated quantitative goals. This is a significant shift towards prioritizing personal meaning rather than easily accessible quantified goal measures.

5.4 Goal-Oriented Work: Qualitative Goals

Measures of motivation are critical, as they include not just scales but also workbooks and exercise sheets for goal-oriented work. Developed as a therapeutic approach, goal-oriented work includes activities in which the focus is on the patient's goals so that they are thoroughly explored [46]. Several papers employed tools or methods inspired by goal-oriented work such as Personal Project Analysis [146, 147] used by Brotman et al. [27], motivational interview [212] used by Chen et al. [35], or interviews with health experts for deeper reflection on goals [137]. This emerging work is important as it highlights a significant shift from the focus on quantitative (concrete and trackable) [62] to qualitative measures of goals that are abstract and not directly trackable [186]. This aligns with the critique of the instrumentalist quality of quantitative tracking and its limited account for the self or user motives and personal meaning of tracked data [207]. Future research directions could benefit for richer use of goal-oriented work methods for more in depth exploration of user needs to inform the design of technologies for behavior change. Exciting opportunities in this space could focus on supporting users to engage with such methods through novel interface designs.

5.5 Stakeholders: WEIRD, Non-Inclusive Gender Language

The importance of representing the different stakeholders in research [144] cannot be over-emphasized, particularly when engaging with technologies for behavior change in-the-wild as their impact may differ across user groups. Consistent with prior research, we find that the ratio of Western to non-Western sites for user studies (4 Western to 1 non-Western) is slightly higher than Linxen et al. [144] (3 Western to 1 non-Western). Surprisingly, a large number of papers do not disclose where the study site is (38%). We were surprised that the demographic of older adults included in studies was present in 28.9% of the papers, when, for example, in the US the older adult population is 16.8% of the total population [249]. We see more opportunities to study children, which in our sample accounted for 10% of the data, compared to the US census where they are 22% of the population [250]. With respect to vulnerable users with diagnosed physical or mental health conditions, a surprising outcome was the prevalence of a few conditions such as diabetes, depression, or anxiety. Future work can address these limitations by exploring additional chronic conditions with a large societal impact.

Regarding gender, we see a lot more need to be inclusive of non-binary individuals, with only 11.7% of the paper including any participants identifying as non-binary. Perhaps more concerning, 78% of the papers had non-inclusive gender language, referring to the sex of participants, when in many cases, only the gender might be relevant. With HCI researchers advocating and explaining how to use inclusive language in surveys we see the language as an immediate area for improvement [237].

5.6 Mobile Apps & Wearables Deployed in Everyday Life, 1-4 Weeks

Our outcomes show that mobile apps are by far the most common type of technology deployed in-the-wild (43.3%), followed by wearables (16.1%). Apps are also the most common technology across all domains, while browser extensions are most used in digital wellbeing. The latter finding differs slightly from previous work, where mobile apps were the most common technology for digital wellbeing tools [210]. Technologies for behavior change were also deployed overwhelmingly under one month (2-31 days) (70% of those deployed in-the-wild). The number of deployments between 1-3 months was roughly similar to those of 1-2 weeks or 2-3 weeks (25% of deployments). Most deployments were in everyday life settings, with less use of specific sites such as school or work, given the goal domains for learning and productivity. Despite the nature of behavior change being longitudinal, taking months or year [206], the current studies are largely limited in length. We argue for more longitudinal deployments that build our understanding of how to support behavior change longitudinally.

5.7 Artifact Contribution vs Theoretical Contribution

Through the lens of HCI contributions [259], the last decade of work in this space has been dominated by artifact design and development (over 60% of papers), through a rich body of empirical research. The large presence of artifact contribution in our findings contrasts sharply with the much narrower theoretical engagement, reflected by limited theoretical contributions and limited theoretically informed design. Such findings align with the acknowledged theoretical gap both in specific domains such as digital wellbeing [210], and more broadly in HCI [93]. While our findings confirm this gap, they also shed light on the tension between the breadth and depth of theoretical engagement, with most papers merely mentioning theories (75%) instead of using them (42%). Digital wellbeing domain stands out with the highest number of used theories (47%), and the highest number of interventions which were deployed, including most diverse set of interventions. A somehow lower 39% of papers using theories for the design of digital wellbeing tools has been recently reported [210]. Given the amount of theory mentioned in papers, one surprising fact is that in light of this, very few papers made a contribution back to theory (5%).

Despite almost three times a higher number of different theories being mentioned rather than used, the number of most commonly used theories is less than a dozen, most of which being among those used by interventions identified as most effective. We would like to see more robust engagement with theories, for instance, through concentrated interdisciplinary efforts to operationalize them so that they can be more easily accessed by HCI researchers, and used to inform design. Rather than extending further the long list of theories that HCI research has rather superficially engaged with, we suggest deeper engagement with a smaller set of theories, such as the ones that informed the design of most effective interventions.

5.8 Reflection on Most Effective Interventions

We now reflect on the most effective technology-based interventions highlighting three design principles to inform technologies focused on goals for behavior change. See Section 4 in Supplementary material for detailed structured descriptions of the five most effective interventions.

Ensuring strong theoretical rationale grounded in one or more theories, whose trade-offs are carefully considered and sensitively addressed through design. For example, in their design of the DStress app supporting difficulty level adaptive goals for two domains: physical exercise and wellbeing, Konrad et al. [120] explored both goal-setting theory and self-determination theory, addressing one tension that may account for poor compliance with health intervention, namely difficult goals are better as they require more effort, versus easy goals are better as they can be completed with less effort which strengthens confidence in approaching future goals. Another example is the GoldenTime app [194], whose design integrates self-regulation theory and behavioral economics theory to reinforce limiting problematic phone usage through timeboxing and reward/loss of micro-financial incentives. Their findings confirm the loss aversion cognitive bias through the significantly less problematic screen time for the loss group. One other example is a cooperative tracker for physical activity [208] whose design draws from a review of HCI work on social features of trackers and the impact of proximity on co-workers' bonding.

Supporting awareness of deep motives for qualitative goals, and goal setting for self-set goals or social goals. Most effective interventions aim to support reflection on users' deep motives for qualitative goals, *prior* to goal setting. For example, Lee et al. [137] interviewed health experts on how to support physical activity and personalization of physical activity plans, and leveraged therapists' approaches, similar to goal-oriented work [27]. Lee et al. [137] showed the importance of supporting users' reflection on deeper motives in order to articulate their goals, as a critical initial step prior to goal setting and tracking. Indeed, users who engaged with reflective questions set ambitious goals and overachieved, by walking significantly more daily steps than those without reflective questions [137]. For goal setting, the Habito [86] app supports self-set goals, whose evaluation shows significantly increased walking distance, compared to app-provided goals. While most of the effective interventions focused on individual goals, an excellent illustration of the less explored social goals is the PCTF wearable system [208] which leverages social support through cooperative rather than competitive exercising, and pre-existing social relationships among pairs of co-located coworkers, as they engage in cooperative, daily step goal setting and sharing of tracked data.

Scaffolding reflection on tracked goals for behavior change. Most effective interventions not just report tracked data but explicitly scaffold reflection on it to support behavior change. For instance, the Habito app contextualizes physical activity with data on visited locations and highlights interesting patterns, such as intense activity in one location in order to prompt reflection [86]. The GoldenTime app provided notifications on failure to gain coins, or deduction of coins when users fail to regulate their screen time, which prompted them to reflect, and reframe their problematic phone use as incurring a price, materialized by the coins as phone usage fee [194].

To summarize, while the current focus of HCI research on goals for behavior change is strongly skewed towards single-domain, individual, low intrinsic, and quantitative goals, we suggest extending this to include also multiple-domain, social, high intrinsic, and qualitative goals. We also suggest more inclusive stakeholders from domain experts to people living with diagnosed conditions, both children and senior adults. Finally, we highlight the limited theoretical underpinning of technology-based interventions, despite the extensive artifact contributions, most of which were evaluated in-the-wild.

6 Limitations and Future Work

We focused on ACM DL because, according to its description, it is "the world's most comprehensive database of full-text articles and bibliographic literature covering computing and information technology" [18]. ACM DL is also the most commonly used database in HCI reported reviews [60, 68, 220]. Limitations related to focusing exclusively on ACM DL pertain to missing other relevant publications in databases such as PubMed. As an initial exploration of our research questions, we argue that using ACM DL as a sole database, as done also by other HCI reviews [60, 211, 225, 240, 244, 265] is appropriate as it aligns best with our HCI centric research questions. Future work should extend the search to other databases, such as PubMed or Electrical and Electronics Engineers (IEEE), and will require additional resources beyond those available to us.

7 Conclusion

In this systematic review of 180 goal-related behavior change papers within SIGCHI literature, we summarized recent trends, such as a dominant focus on single-domain, individual goals, especially those related to physical health and wellness that are not intrinsic to users. We also reported on a variety of goal measures, the emergence of goal-oriented work, and the effectiveness of different intervention techniques for different domains of goals. We pointed out concerning trends about limited inclusiveness of children's goals, non-western contexts, and the limited use of theoretically-informed design of goal technologies. We further expanded the design space for behavior change goal-related research and urge HCI researchers to: *design for* multi-domain, highly intrinsic, and social goals and *design technology-based interventions* better grounded in theoretical frameworks.

Acknowledgments

The last author's work was partly supported by Digital health project, funded by EPSRC, EP/W003228/1.

8 Note

Within References, the 180 reviewed papers are marked with an asterisk (*), and from these, the 37 papers that were reviewed for interventions' effectiveness are marked with two asterisks (**).

References

[1] * Parastoo Abtahi, Victoria Ding, Anna C Yang, Tommy Bruzzese, Alyssa B Romanos, Elizabeth L Murnane, Sean Follmer, and James A Landay. 2020. Understanding physical practices and the role of technology in manual self-tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.

[2] * Alexander T Adams, Jean Costa, Malte F Jung, and Tanzeem Choudhury. 2015. Mindless computing: designing technologies to subtly influence behavior. In *Proceedings of the 2015 ACM international joint conference on pervasive and*

ubiquitous computing. Association for Computing Machinery, New York, NY, USA, 719–730.

[3] * Elena Agapie, Patricia A Areán, Gary Hsieh, and Sean A Munson. 2022. A longitudinal goal setting model for addressing complex personal problems in mental health. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.

[4] ** Elena Agapie, Daniel Avrahami, and Jennifer Marlow. 2016. Staying the course: System-driven lapse management for supporting behavior change. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1072–1083.

[5] * Elena Agapie, Bonnie Chinh, Laura R Pina, Diana Oviedo, Molly C Welsh, Gary Hsieh, and Sean Munson. 2018. Crowdsourcing Exercise plans aligned with expert guidelines and everyday constraints. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[6] * Elena Agapie, Lucas Colusso, Sean A Munson, and Gary Hsieh. 2016. Plansourcing: Generating behavior change plans with friends and crowds. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. Association for Computing Machinery, New York, NY, USA, 119–133.

[7] * Aino Ahtinen, Eeva Andrejeff, Christopher Harris, and Kaisa Väänänen. 2017. Let's walk at work: persuasion through the brainwolk walking meeting app. In *Proceedings of the 21st International Academic Mindtrek Conference*. Association for Computing Machinery, New York, NY, USA, 73–82.

[8] * Joelle Alcaidinho, Giancarlo Valentin, Stephanie Tai, Brian Nguyen, Krista Sanders, Melody Jackson, Eric Gilbert, and Thad Starner. 2015. Leveraging mobile technology to increase the permanent adoption of shelter dogs. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 463–469.

[9] Mariam Alenazi and Corina Sas. 2023. Evaluating Budgeting Apps: Limited Support for Budgeting Compared to Tracking. In *36th International BCS Human-Computer Interaction Conference*. BCS Learning & Development, BCS, The Chartered Institute for IT, 1–12.

[10] Susan Alisat, Joan E Norris, Michael W Pratt, M Kyle Matsuba, and Dan P McAdams. 2014. Caring for the earth: Generativity as a mediator for the prediction of environmental narratives from identity among activists and nonactivists. *Identity* 14, 3 (2014), 177–194.

[11] Sultan Almoallim, Corina Sas, et al. 2022. Toward research-informed design implications for interventions limiting smartphone use: functionalities review of digital well-being apps. *JMIR formative research* 6, 4 (2022), e31730.

[12] * Najla Almutari and Rita Orji. 2021. Culture and Health Belief Model: Exploring the determinants of physical activity among Saudi adults and the moderating effects of age and gender. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 138–146.

[13] ** Deemah Alqahtani, Caroline Jay, and Markel Vigo. 2020. The effect of goal moderation on the achievement and satisfaction of physical activity goals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–18.

[14] * Deemah Alqahtani, Caroline Jay, and Markel Vigo. 2020. The role of uncertainty as a facilitator to reflection in self-tracking. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 1807–1818.

[15] * Swamy Ananthanarayan, Nathan Lapinski, Katie Siek, and Michael Eisenberg. 2014. Towards the crafting of personal health technologies. In *Proceedings of the 2014 conference on Designing interactive systems*. Association for Computing Machinery, New York, NY, USA, 587–596.

[16] * Swamy Ananthanarayan, Katie Siek, and Michael Eisenberg. 2016. A craft approach to health awareness in children. In *Proceedings of the 2016 ACM conference on designing interactive systems*. Association for Computing Machinery, New York, NY, USA, 724–735.

[17] * Ian Arawjo, Ariam Mogos, Steven J Jackson, Tapan Parikh, and Kentaro Toyama. 2019. Computing education for intercultural learning: Lessons from the Nairobi play project. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[18] Association for Computing Machinery. 2024. *ACM Digital Library*. Association for Computing Machinery. https://www.acm.org/publications/digital-library

[19] James T Austin and Jeffrey B Vancouver. 1996. Goal constructs in psychology: Structure, process, and content. *Psychological bulletin* 120, 3 (1996), 338.

[20] * Amid Ayobi, Paul Marshall, Anna L Cox, and Yunan Chen. 2017. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 6889–6901.

[21] Albert Bandura. 1991. Social cognitive theory of self-regulation. *Organizational behavior and human decision processes* 50, 2 (1991), 248–287.

[22] Albert Bandura et al. 1986. Social foundations of thought and action. *Englewood Cliffs, NJ* 1986, 23-28 (1986), 2.

[23] * Andrea M Barbarin, Laura R Saslow, Mark S Ackerman, and Tiffany C Veinot. 2018. Toward health information technology that supports overweight/obese women in addressing emotion-and stress-related eating. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[24] * Fadi Botros, Charles Perin, Bon Adriel Aseniero, and Sheelagh Carpendale. 2016. Go and grow: Mapping personal data to a living plant. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. Association for Computing Machinery, New York, NY, USA, 112–119.

[25] Veronika Brandstätter and Marie Hennecke. 2018. Goals. In *Motivation and action*. Springer, Cham, Switzerland, 453–484.

[26] * Nathalie Bressa, Jo Vermeulen, and Wesley Willett. 2022. Data Every Day: Designing and Living with Personal Situated Visualizations. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18.

[27] * Ryan Brotman, Winslow Burleson, Jodi Forlizzi, William Heywood, and Jisoo Lee. 2015. Building change: Constructive design of smart domestic environments for goal achievement. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3083–3092.

[28] Mohamad Adam Bujang and Nurakmal Baharum. 2017. Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiology, biostatistics, and public health* 14, 2 (2017), e12267–1–e12267–10.

[29] * Eleanor R Burgess, Elizabeth Kaziunas, and Maia Jacobs. 2022. Care Frictions: A Critical Reframing of Patient Noncompliance in Health Technology Design. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[30] ** Jessica R Cauchard, Jeremy Frey, Octavia Zahrt, Krister Johnson, Alia Crum, and James A Landay. 2019. The positive impact of push vs pull progress feedback: a 6-week activity tracking study in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–23.

[31] ** Samantha WT Chan, Thisum Buddhika, Haimo Zhang, and Suranga Nanayakkara. 2019. Prospecfit: In situ evaluation of digital prospective memory training for older adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–20.

[32] ** Kerry Shih-Ping Chang, Catalina M Danis, and Robert G Farrell. 2014. Lunch line: using public displays and mobile devices to encourage healthy eating in an organization. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 823–834.

[33] * Beenish Moalla Chaudhry, Dipanwita Dasgupta, and Nitesh Chawla. 2022. Formative evaluation of a tablet application to support goal-oriented care in community-dwelling older adults. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–21.

[34] ** Beenish M Chaudhry, Christopher Schaefbauer, Ben Jelen, Katie A Siek, and Kay Connelly. 2016. Evaluation of a food portion size estimation interface for a varying literacy population. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5645–5657.

[35] * Tianying Chen, Kristy Zhang, Robert E Kraut, and Laura Dabbish. 2021. Scaffolding the online peer-support experience: novice supporters' strategies and challenges. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–30.

[36] * Ananta Chowdhury and Andrea Bunt. 2023. Co-Designing with Early Adolescents: Understanding Perceptions of and Design Considerations for Tech-Based Mediation Strategies that Promote Technology Disengagement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[37] * Tee Chuanromanee and Ronald Metoyer. 2023. Understanding gender transition tracking habits and technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[38] Ada S Chulef, Stephen J Read, and David A Walsh. 2001. A hierarchical taxonomy of human goals. *Motivation and Emotion* 25 (2001), 191–232.

[39] * Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When personal tracking becomes social: Examining the use of Instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1674–1687.

[40] * Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A Munson. 2019. Identifying and planning for individualized change: Patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 1 (2019), 1–27.

[41] * Franceli L Cibrian, Kimberley D Lakes, Arya Tavakoulnia, Kayla Guzman, Sabrina Schuck, and Gillian R Hayes. 2020. Supporting self-regulation of children with ADHD using wearables: tensions and design challenges. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[42] * Ana Ciocarlan, Judith Masthoff, and Nir Oren. 2018. Kindness is contagious: study into exploring engagement and adapting persuasive games for wellbeing. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 311–319.

[43] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[44] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 457–466.

[45] Sunny Consolvo, David W McDonald, and James A Landay. 2009. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 405–414.

[46] Mick Cooper and Duncan Law. 2018. *Working with Goals in Psychotherapy and Counselling*. Oxford University Press, Oxford University. doi:10.1093/med-psych/9780198793687.001.0001

[47] * Saskia Coulson, Mel Woods, Michelle Scott, Drew Hemment, and Mara Balestrini. 2018. Stop the noise! enhancing meaningfulness in participatory sensing with community level indicators. In *Proceedings of the 2018 designing interactive systems conference*. Association for Computing Machinery, New York, NY, USA, 1183–1192.

[48] Geert Crombez, Emelien Lauwerier, Liesbet Goubert, and Stefaan Van Damme. 2016. Goal pursuit in individuals with chronic pain: a personal project analysis. *Frontiers in Psychology* 7 (2016), 966.

[49] * Nediyana Daskalova, Eindra Kyi, Kevin Ouyang, Arthur Borem, Sally Chen, Sung Hyun Park, Nicole Nugent, and Jeff Huang. 2021. Self-e: Smartphone-supported guidance for customizable self-experimentation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[50] ** Nediyana Daskalova, Bongshin Lee, Jeff Huang, Chester Ni, and Jessica Lundin. 2018. Investigating the effectiveness of cohort-based sleep recommendations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–19.

[51] Claudia Daudén Roquet and Corina Sas. 2021. Interoceptive interaction: An embodied metaphor inspired approach to designing for meditation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[52] Claudia Daudén Roquet, Corina Sas, and Dominic Potts. 2023. Exploring Anima: a brain–computer interface for peripheral materialization of mindfulness states during mandala coloring. *Human–Computer Interaction* 38, 5-6 (2023), 259–299.

[53] ** Katie Davis, Petr Slovak, Rotem Landesman, Caroline Pitt, Abdullatif Ghajar, Jessica Lee Schleider, Saba Kawas, Andrea Guadalupe Perez Portillo, and Nicole S Kuhn. 2023. Supporting Teens' Intentional Social Media Use Through Interaction Design: An exploratory proof-of-concept study. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. Association for Computing Machinery, New York, NY, USA, 322–334.

[54] * Alynne De Haan, Daphne Menheere, Steven Vos, and Carine Lallemand. 2021. Aesthetic of friction for exercising motivation: a prototyping journey. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 1056–1067.

[55] Margarida Pedroso de Lima, Isabel Albuquerque, Paulo Jorge Martins, and António-José Gonzalez. 2023. Personal Projects Analysis as an idiographic approach in psychotherapy: an exploratory study. *Research in Psychotherapy: Psychopathology, Process, and Outcome* 26, 1 (2023), 1–11.

[56] Edward L Deci, Richard Koestner, and Richard M Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin* 125, 6 (1999), 627.

[57] Edward L Deci and Richard M Ryan. 2008. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne* 49, 3 (2008), 182.

[58] * Teresa Denefleh, Arne Berger, Albrecht Kurze, Andreas Bischof, and Christopher Frauenberger. 2019. Sensorstation: Exploring simple sensor data in the context of a shared apartment. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 683–695.

[59] * Pooja M Desai, Matthew E Levine, David J Albers, and Lena Mamykina. 2018. Pictures worth a thousand words: Reflections on visualizing personal blood glucose forecasts for individuals with type 2 diabetes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[60] Tawanna R. Dillahunt, Xinyi Wang, Earnest Wheeler, Hao Fei Cheng, Brent Hecht, and Haiyi Zhu. 2017. The Sharing Economy in Computing: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 38 (Dec. 2017), 26 pages. doi:10.1145/3134673

[61] * Julie Doyle, Emma Murphy, Janneke Kuiper, Suzanne Smith, Caoimhe Hannigan, An Jacobs, and John Dinsmore. 2019. Managing multimorbidity: identifying design requirements for a digital self-management tool to support older adults with multiple chronic conditions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[62] Tina Ekhtiar, Armağan Karahanoğlu, Rúben Gouveia, and Geke Ludden. 2023. Goals for goal setting: a scoping review on personal informatics. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 2625–2641.

[63] * Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*. Association for Computing Machinery, New York, NY, USA, 667–676.

[64] * Daniel A Epstein, Alan Borning, and James Fogarty. 2013. Fine-grained sharing of sensed physical activity: A value sensitive approach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 489–498.

[65] * Daniel A Epstein, Jennifer H Kang, Laura R Pina, James Fogarty, and Sean A Munson. 2016. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 829–840.

[66] * Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean Munson. 2017. Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, New York, NY, USA, 6876–6888.

[67] * Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 731–742.

[68] Daniel A Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, et al. 2020. Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–38.

[69] ** Thomas Erickson, Ming Li, Younghun Kim, Ajay Deshpande, Sambit Sahu, Tian Chao, Piyawadee Sukaviriya, and Milind Naphade. 2013. The dubuque electricity portal: evaluation of a city-scale residential electricity consumption feedback system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1203–1212.

[70] * Sindhu Kiranmai Ernala, Kathan H Kashiparekh, Amir Bolous, Asra Ali, John M Kane, Michael L Birnbaum, and Munmun De Choudhury. 2021. A social media study on mental health status transitions surrounding psychiatric hospitalizations. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–32.

[71] * Jordan Eschler, Eleanor R Burgess, Madhu Reddy, and David C Mohr. 2020. Emergent self-regulation practices in technology and social media use of individuals living with depression. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[72] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[73] * Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. 2018. People like me: Designing for reflection on aggregate cohort data in personal informatics systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.

[74] Ayelet Fishbach, Janina Steinmetz, and Yanping Tu. 2016. Motivation in a social context: coordinating personal and shared goal pursuits with others. *Advances in motivation science* 3 (2016), 35–79.

[75] Ayelet Fishbach and Kaitlin Woolley. 2022. The structure of intrinsic motivation. *Annual Review of Organizational Psychology and Organizational Behavior* 9, 1 (2022), 339–363.

[76] * Jaimie Lee Freeman and Amanda Nicole Curtis. 2023. Putting the Self in Self-Tracking: The Value of a Co-Designed 'How Might You'Self-Tracking Guide for Teenagers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[77] * Thomas Fritz, Elaine M Huang, Gail C Murphy, and Thomas Zimmermann. 2014. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 487–496.

[78] * Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells,

Jonathan N. Tobin, Esteban G. Tabak, et al. 2021. From reflection to action: combining machine learning with expert knowledge for nutrition goal recommendations. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. ACM New York, NY, USA, New York, NY, USA, 1–17.

[79] ** Anirudh Ganesh, Chinenye Ndulue, and Rita Orji. 2023. Tailoring a persuasive game to promote secure smartphone behaviour. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18.

[80] Gerald Gartlehner, Richard A Hansen, Daniel Nissman, Kathleen N Lohr, and Timothy S Carey. 2010. Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. *AHRQ publication no. 06-0046. 2006. Rockville* Technical review 12 (2010), 6–46.

[81] * Hüseyin Uğur Genç, Duru Erdem, Çağla Yıldırım, and Aykut Coskun. 2022. Mind the Whisper: Enriching Collocated Social Interactions in Public Places through Audio Narratives. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. ACM New York, NY, USA, New York, NY, USA, 1428–1440.

[82] * Eva Geurts, Fanny Van Geel, Peter Feys, and Karin Coninx. 2019. WalkWithMe: personalized goal setting and coaching for walking in people with multiple sclerosis. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 51–60.

[83] Peter M Gollwitzer. 1999. Implementation intentions: strong effects of simple plans. *American psychologist* 54, 7 (1999), 493.

[84] * Daniel Gooch, Blaine A Price, Anna Klis-Davies, and Julie Webb. 2021. A design exploration of health-related community displays. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–22.

[85] * Nanna Gorm and Irina Shklovski. 2017. Participant driven photo elicitation for understanding activity tracking: Benefits and limitations. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. Association for Computing Machinery, New York, NY, USA, 1350–1361.

[86] ** Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How do we engage with activity trackers? A longitudinal study of Habito. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 1305–1316.

[87] * Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2018. Activity tracking in vivo. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[88] * Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A Munson, and Marc Hassenzahl. 2016. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 144–155.

[89] Lou Grimal, Ines Di Loreto, and Nadège Troussier. 2023. Value misalignments in interactions: an opportunity for sustainable HCI. In *IHM'23: Proceedings of the 34th Conference on l'Interaction Humain-Machine*. l'Interaction Humain-Machine, Troyes, France, 2–14.

[90] * Laura Grönewald, Julian Weiblen, Matthias Laschke, Lara Christoforakos, and Marc Hassenzahl. 2023. Sustainability by design. How to encourage users to choose energy-saving programs and settings when washing laundry. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[91] * Xinning Gui, Yu Chen, Clara Caldeira, Dan Xiao, and Yunan Chen. 2017. When fitness meets social networks: Investigating fitness tracking and social practices on werun. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1647–1659.

[92] Jessica Gurevitch, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. 2018. Meta-analysis and the science of research synthesis. *Nature* 555, 7695 (2018), 175–182.

[93] Eric B Hekler, Predrag Klasnja, Jon E Froehlich, and Matthew P Buman. 2013. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3307–3316.

[94] Brenda E Hogan, Wolfgang Linden, and Bahman Najarian. 2002. Social support interventions: do they work? *Clinical psychology review* 22, 3 (2002), 381–440.

[95] ** Victoria Hollis, Artie Konrad, and Steve Whittaker. 2015. Change of heart: emotion tracking to promote behavior change. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2643–2652.

[96] ** Nanna Inie, Bjørn Hjorth Westh, John Henrik Muller, and Mircea Filip Lungu. 2023. Challenges and Opportunities of Using Redirection of Activity for Self-Regulation Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–19.

[97] ** Ellen Isaacs, Artie Konrad, Alan Walendowski, Thomas Lennig, Victoria Hollis, and Steve Whittaker. 2013. Echoes from the past: how technology mediated reflection improves well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1071–1080.

[98] * Arne Jansen, Maarten Van Mechelen, and Karin Slegers. 2017. Personas and behavioral theories: A case study using self-determination theory to construct overweight personas. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2127–2136.

[99] * Matthew Jörke, Yasaman S Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. 2023. Pearl: A technology probe for machine-assisted reflection on personal data. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 902–918.

[100] * Somi Ju, Kyung-Ryong Lee, Subin Kim, and Young-Woo Park. 2019. Bookly: An interactive everyday artifact showing the time of physically accumulated reading activity. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–8.

[101] ** Gyuwon Jung, Jio Oh, Youjin Jung, Juho Sun, Ha-Kyung Kong, and Uichin Lee. 2021. "Good Enough!": Flexible Goal Achievement with Margin-based Outcome Evaluation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[102] * Heekyoung Jung. 2020. In Search of Forms for Evocative and Generative Reflection: Exploratory Studies and a Design Proposal. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[103] * Adela Kapuscinska, Payal M Bhujwala, Melissa Kalarchian, and Jessica Hammer. 2021. A Socio-Ecological Approach to Activity Games for Girls. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–28.

[104] * Kasper Karlgren and Donald Mcmillan. 2023. Sleep Planning with Awari: Uncovering the Materiality of Body Rhythms using Research through Design. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, New York, NY, USA, 1–17.

[105] * Dmitri S Katz, Blaine A Price, Simon Holland, and Nicholas Sheep Dalton. 2018. Designing for diabetes decision support systems with fluid contextual reasoning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[106] * Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 629–641.

[107] * Inyeop Kim, Hwarang Goh, Nematjon Narziev, Youngtae Noh, and Uichin Lee. 2020. Understanding user contexts and coping strategies for context-aware phone distraction management system design. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–33.

[108] * Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's FOCUS: mitigating mobile phone use in college classrooms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–29.

[109] * Inyeop Kim, Minsam Ko, Joonyoung Park, Sung Wook Moon, Gyuwon Jung, Youn-kyung Lim, and Uichin Lee. 2022. Social-Spiritual Face: Designing Social Reading Support for Spiritual Well-being. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.

[110] ** Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology supported behavior restriction for mitigating self-interruptions in multi-device environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[111] ** Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. Goalkeeper: Exploring interaction lockout mechanisms for regulating smartphone use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–29.

[112] * Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout task intervention for discouraging smartphone app use. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[113] * Yoojung Kim, Hee-Tae Jung, Joonwoo Park, Yangsoo Kim, Nathan Ramasarma, Paolo Bonato, Eun Kyoung Choe, and Sunghoon Ivan Lee. 2019. Towards the design of a ring sensor-based mHealth system to achieve optimal motor function in stroke survivors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–26.

[114] * Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–28.

[115] * Jesper Kjeldskov, Mikael B Skov, Jeni Paay, and Rahuvaran Pathmanathan. 2012. Using mobile phones to support sustainability: a field study of residential electricity consumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2347–2356.

[116] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 3063–3072.

[117] Predrag Klasnja, Eric B Hekler, Elizabeth V Korinek, John Harlow, and Sonali R Mishra. 2017. Toward usable evidence: optimizing knowledge accumulation in HCI research on health behavior change. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 3071–3082.

[118] * Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for workplace reflection: a chat and voice-based conversational agent. In *Proceedings of the 2018 designing interactive systems conference*. Association for Computing Machinery, New York, NY, USA, 881–894.

[119] * Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: a conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–26.

[120] ** Artie Konrad, Victoria Bellotti, Nicole Crenshaw, Simon Tucker, Les Nelson, Honglu Du, Peter Pirolli, and Steve Whittaker. 2015. Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3829–3838.

[121] * Rachel Kornfield, David C Mohr, Rachel Ranney, Emily G Lattie, Jonah Meyerhoff, Joseph J Williams, and Madhu Reddy. 2022. Involving Crowdworkers with lived experience in content-development for push-based digital mental health tools: lessons learned from crowdsourcing mental health messages. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW1 (2022), 1–30.

[122] ** Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. 2019. Conservation of procrastination: Do productivity interventions save time or just redistribute it?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[123] ** Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. 2018. Rotating online behavior change interventions increases effectiveness but also increases attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

[124] * Sneha R Krishna Kumaran, Yue Yin, and Brian P Bailey. 2021. Plan early, revise more: effects of goal setting and perceived role of the feedback provider on feedback seeking behavior. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–22.

[125] Arie W Kruglanski, James Y Shah, Ayelet Fishbach, Ron Friedman, Woo Young Chun, and David Sleeth-Keppler. 2018. A theory of goal systems. *The motivated mind* 34 (2018), 207–250.

[126] * Kaylee Payne Kruzan, Ada Ng, Colleen Stiles-Shields, Emily G Lattie, David C Mohr, and Madhu Reddy. 2023. The perceived utility of smartphone and wearable sensor data in digital self-tracking technologies for mental health. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[127] Franki Y H Kung and Abigail A Scholer. 2021. Moving Beyond Two Goals: An Integrative Review and Framework for the Study of Multiple Goals. *Pers. Soc. Psychol. Rev.* 25, 2 (May 2021), 130–158.

[128] * Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. 2019. Exploring the state-of-receptivity for mHealth interventions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[129] Emily R Lai. 2011. Motivation: A literature review. *Person Research's Report* 6 (2011), 40–41.

[130] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* 4 (2013), 863.

[131] * Rithika Lakshminarayanan, Alexandra Canori, Aditya Ponnada, Melissa Nunn, Mary Schmidt Read, Shivayogi V Hiremath, and Stephen Intille. 2022. Exploring Opportunities to Improve Physical Activity in Individuals with Spinal Cord Injury Using Context-Aware Messaging. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[132] * Laura Lascau, Sandy JJ Gould, Anna L Cox, Elizaveta Karmannaya, and Duncan P Brumby. 2019. Monotasking or multitasking: Designing for crowdworkers' preferences. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[133] * Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H Nguyen. 2015. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 635–646.

[134] ** 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 *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 6837–6849.

[135] * Jong Ho Lee, Jessica Schroeder, and Daniel A Epstein. 2021. Understanding and supporting self-tracking app selection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–25.

[136] * Kwangyoung Lee and Hwajung Hong. 2018. MindNavigator: Exploring the stress and self-interventions for mental wellness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[137] ** Min Kyung Lee, Junsung Kim, Jodi Forlizzi, and Sara Kiesler. 2015. Personalization revisited: a reflective approach helps people better personalize health services and motivates them to increase physical activity. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 743–754.

[138] ** Matthew L Lee and Anind K Dey. 2014. Real-time feedback for improving medication taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2259–2268.

[139] * Roberto Legaspi, Wenzhen Xu, Tatsuya Konishi, Shinya Wada, and Yuichi Ishikawa. 2022. Multidimensional analysis of sense of agency during goal pursuit. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 34–47.

[140] * Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. " Don't Whip Me With Your Games" Investigating" Bottom-Up" Gamification. In *Proceedings of the 2016 chi conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2026–2037.

[141] * Yu Liang, Aditya Ponnada, Paul Lamere, and Nediyana Daskalova. 2023. Enabling goal-focused exploration of podcasts in interactive recommender systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 142–155.

[142] * Yu Liang and Martijn C Willemsen. 2019. Personalized recommendations for music genre exploration. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 276–284.

[143] * Brian Y Lim, Judy Kay, and Weilong Liu. 2019. How does a nation walk? Interpreting large-scale step count activity with weekly streak patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–46.

[144] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How weird is CHI?. In *Proceedings of the 2021 chi conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[145] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 143, 14 pages. doi:10.1145/3411764.3445488

[146] Brian R Little. 1983. Personal projects: A rationale and method for investigation. *Environment and behavior* 15, 3 (1983), 273–309.

[147] Brian R Little. 2014. Well-doing: Personal projects and the quality of lives. *Theory and Research in Education* 12, 3 (2014), 329–346.

[148] Edwin A Locke and Gary P Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist* 57, 9 (2002), 705.

[149] Edwin A Locke and Gary P Latham. 2019. The development of goal setting theory: A half century retrospective. *Motivation Science* 5, 2 (2019), 93.

[150] Edwin A Locke, Karyll N Shaw, Lise M Saari, and Gary P Latham. 1981. Goal setting and task performance: 1969–1980. *Psychological bulletin* 90, 1 (1981), 125.

[151] Sruzan Lolla and Corina Sas. 2023. Evaluating Mobile Apps Targeting Personal Goals. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7.

[152] * Xi Lu, Yunan Chen, and Daniel A Epstein. 2021. A model of socially sustained self-tracking for food and diet. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–32.

[153] * Xi Lu, Edison Thomaz, and Daniel A Epstein. 2022. Understanding People's Perceptions of Approaches to Semi-Automated Dietary Monitoring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.

[154] * Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y Lim. 2018. TableChat: mobile food journaling to facilitate family support for healthy eating. *Proceedings of*

the ACM on Human-Computer Interaction 2, CSCW (2018), 1–28.

[155] * Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[156] * John R Lund and Jason Wiese. 2021. Less is more: exploring support for time management planning. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 392–405.

[157] ** Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted' Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[158] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *proceedings of the 2019 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–18.

[159] Aaron Lyon, Sean A Munson, Madhu Reddy, Stephen M Schueller, Elena Agapie, Svetlana Yarosh, Alex Dopp, Ulrica von Thiele Schwarz, Gavin Doherty, Andrea K Graham, et al. 2023. Bridging HCI and Implementation Science for Innovation Adoption and Public Health Impact. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7.

[160] * Andrew Macvean and Judy Robertson. 2013. Understanding exergame users' physical activity, motivation and behavior over time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1251–1260.

[161] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 477–486.

[162] Jennifer C Mankoff, Eli Blevis, Alan Borning, Batya Friedman, Susan R Fussell, Jay Hasbrouck, Allison Woodruff, and Phoebe Sengers. 2007. Environmental sustainability and interaction. In *CHI'07 extended abstracts on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2121–2124.

[163] Bess H Marcus, Vanessa C Selby, Raymond S Niaura, and Joseph S Rossi. 1992. Self-efficacy and the stages of exercise behavior change. *Research quarterly for exercise and sport* 63, 1 (1992), 60–66.

[164] * Aqueasha Martin-Hammond and Tanjala S Purnell. 2022. Bridging community, history, and culture in personal informatics tools: Insights from an existing community-based heart health intervention for Black Americans. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–23.

[165] Christopher M Masi, Hsi-Yuan Chen, Louise C Hawkley, and John T Cacioppo. 2011. A meta-analysis of interventions to reduce loneliness. *Personality and social psychology review* 15, 3 (2011), 219–266.

[166] ** Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. 2013. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on human factors in Computing systems*. Association for Computing Machinery, New York, NY, USA, 371–380.

[167] * Yelena Mejova and Kyriaki Kalimeri. 2019. Effect of values and technology use on exercise: Implications for personalized behavior change interventions. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 36–45.

[168] * Andre N Meyer, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. 2017. Design recommendations for self-monitoring in the workplace: Studies in software development. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[169] * onah Meyerhoff, Rachel Kornfield, David C Mohr, and Madhu Reddy. 2022. Meeting young adults' social support needs across the health behavior change journey: implications for digital mental health tools. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–33.

[170] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine* 46, 1 (2013), 81–95.

[171] Susan Michie, Maartje M Van Stralen, and Robert West. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science* 6 (2011), 1–12.

[172] * Varun Mishra, Florian Künzler, Jan-Niklas Kramer, Elgar Fleisch, Tobias Kowatsch, and David Kotz. 2021. Detecting receptivity for mHealth interventions in the natural environment. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 2 (2021), 1–24.

[173] ** Elliot G Mitchell, Rosa Maimone, Andrea Cassells, Jonathan N Tobin, Patricia Davidson, Arlene M Smaldone, and Lena Mamykina. 2021. Automated vs. human health coaching: exploring participant and practitioner experiences. *Proceedings of the ACM on human-computer interaction* 5, CSCW1 (2021), 1–37.

[174] ** Asuka Miyake, Masami Takahashi, Ryo Hashimoto, and Momoko Nakatani. 2021. Stepup forecast: predicting future to promote walking. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. Association for Computing Machinery, New York, NY, USA, 1–12.

[175] * Jimmy Moore, Pascal Goffin, Jason Wiese, and Miriah Meyer. 2021. An interview method for engaging personal data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–28.

[176] * Alistair Morrison and Viktor Bakayov. 2017. Stickers for steps: a study of an activity tracking system with face-to-face social engagement. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–10.

[177] * Carol Moser, Sarita Y Schoenebeck, and Paul Resnick. 2019. Impulse buying: Design practices and consumer needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[178] ** Sean A Munson, Erin Krupka, Caroline Richardson, and Paul Resnick. 2015. Effects of public commitments and accountability in a technology-supported physical activity intervention. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1135–1144.

[179] ** Elizabeth L Murnane, Yekaterina S Glazko, Jean Costa, Raymond Yao, Grace Zhao, Paula ML Moya, and James A Landay. 2023. Narrative-Based Visual Feedback to Encourage Sustained Physical Activity: A Field Trial of the WhoIsZuki Mobile Health Platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (2023), 1–36.

[180] Camille Nadal, Caroline Earley, Angel Enrique, Corina Sas, Derek Richards, and Gavin Doherty. 2023. Patient Acceptance of Self-Monitoring on a Smartwatch in a Routine Digital Therapy: A Mixed-Methods Study. *ACM Transactions on Computer-Human Interaction* 31, 1 (2023), 1–50.

[181] Camille Nadal, Corina Sas, and Gavin Doherty. 2020. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. *Journal of Medical Internet Research* 22, 7 (2020), e17256.

[182] * Joshua Newn, Ryan M Kelly, Simon D'Alfonso, and Reeva Lederman. 2022. Examining and promoting explainable recommendations for personal sensing technology acceptance. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.

[183] * Ada Ng, Ashley Marie Walker, Laurie Wakschlag, Nabil Alshurafa, and Madhu Reddy. 2022. Understanding Self-Tracked Data from Bounded Situational Contexts. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 1684–1697.

[184] * Jasmin Niess, Kristina Knaving, Alina Kolb, and Paweł W Woźniak. 2020. Exploring fitness tracker visualisations to avoid rumination. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 1–11.

[185] Jasmin Niess and Paweł W Woźniak. 2018. Supporting meaningful personal fitness: The tracker goal evolution model. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[186] Jasmin Niess and Paweł W. Woźniak. 2018. Supporting Meaningful Personal Fitness: the Tracker Goal Evolution Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173745

[187] ** Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good vibrations: can a digital nudge reduce digital overload?. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services*. Association for Computing Machinery, New York, NY, USA, 1–12.

[188] * Teresa K O'Leary, Dhaval Parmar, Stefan Olafsson, Michael Paasche-Orlow, Timothy Bickmore, and Andrea G Parker. 2022. Community dynamics in technospiritual interventions: lessons learned from a church-based mHealth pilot. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–23.

[189] * Rita Orji, Regan L Mandryk, Julita Vassileva, and Kathrin M Gerling. 2013. Tailoring persuasive health games to gamer type. In *Proceedings of the sigchi conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2467–2476.

[190] * Rita Orji, Lennart E Nacke, and Chrysanne Di Marco. 2017. Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1015–1027.

[191] * Rita Orji, Gustavo F Tondello, and Lennart E Nacke. 2018. Personalizing persuasive strategies in gameful systems to gamification user types. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[192] * Kiemute Oyibo, Ifeoma Adaji, Rita Orji, Babatunde Olabenjo, Mahsa Azizi, and Julita Vassileva. 2018. Perceived persuasive effect of behavior model design in fitness apps. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 219–228.

[193] Matthew J Page, David Moher, and Joanne E McKenzie. 2022. Introduction to PRISMA 2020 and implications for research synthesis methodologists. *Research synthesis methods* 13, 2 (2022), 156–163.

[194] ** Joonyoung Park, Hyunsoo Lee, Sangkeun Park, Kyong-Mee Chung, and Uichin Lee. 2021. Goldentime: exploring system-driven timeboxing and micro-financial incentives for self-regulated phone use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[195] * Joonyoung Park and Uichin Lee. 2023. Understanding disengagement in just-in-time mobile health interventions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–27.

[196] * Gaurav Paruthi, Shriti Raj, Seungjoo Baek, Chuyao Wang, Chuan-che Huang, Yung-Ju Chang, and Mark W Newman. 2018. Heed: exploring the design of situated self-reporting devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.

[197] * Gaurav Paruthi, Shriti Raj, Natalie Colabianchi, Predrag Klasnja, and Mark W Newman. 2018. Finding the sweet spot (S) understanding context to support physical activity plans. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–17.

[198] * Misha Patel and Aisling Ann O'kane. 2015. Contextual influences on the use and non-use of digital technology while exercising at the gym. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2923–2932.

[199] Mark Perry and Jennifer Ferreira. 2018. Moneywork: Practices of use and social interaction around digital and analog money. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 6 (2018), 1–32.

[200] * Svenja Pieritz, Mohammed Khwaja, A Aldo Faisal, and Aleksandar Matic. 2021. Personalised recommendations in mental health apps: the impact of autonomy and data sharing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[201] Laura R Pina, Sang-Wha Sien, Teresa Ward, Jason C Yip, Sean A Munson, James Fogarty, and Julie A Kientz. 2017. From personal informatics to family informatics: Understanding family practices around health monitoring. In *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing*. Association for Computing Machinery, New York, NY, USA, 2300–2315.

[202] * Charlie Pinder, Jo Vermeulen, Benjamin R Cowan, Russell Beale, and Robert J Hendley. 2017. Exploring the feasibility of subliminal priming on smartphones. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, New York, NY, USA, 1–15.

[203] Paul R Pintrich. 1999. The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research* 31, 6 (1999), 459–470.

[204] Joshua R Polanin and Terri D Pigott. 2015. The use of meta-analytic statistical significance testing. *Research Synthesis Methods* 6, 1 (2015), 63–73.

[205] * Ari H Pollack, Uba Backonja, Andrew D Miller, Sonali R Mishra, Maher Khelifi, Logan Kendall, and Wanda Pratt. 2016. Closing the gap: supporting patients' transition to self-management after hospitalization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5324–5336.

[206] James O Prochaska and Wayne F Velicer. 1997. The transtheoretical model of health behavior change. *American journal of health promotion* 12, 1 (1997), 38–48.

[207] Amon Rapp and Maurizio Tirassa. 2017. Know thyself: a theory of the self for personal informatics. *Human–Computer Interaction* 32, 5-6 (2017), 335–380.

[208] ** Xipei Ren, Bin Yu, Yuan Lu, and Aarnout Brombacher. 2018. Exploring cooperative fitness tracking to encourage physical activity among office workers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[209] * Saeyoung Rho, Injung Lee, Hankyung Kim, Jonghyuk Jung, Hyungi Kim, Bong Gwan Jun, and Youn-kyung Lim. 2017. Futureself: what happens when we forecast self-trackers? Future health statuses?. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. Association for Computing Machinery, New York, NY, USA, 637–648.

[210] Alberto Monge Roffarello and Luigi De Russis. 2023. Achieving digital wellbeing through digital self-control tools: A systematic review and meta-analysis. *ACM Transactions on Computer-Human Interaction* 30, 4 (2023), 1–66.

[211] Katja Rogers, Teresa Hirzle, Sukran Karaosmanoglu, Paula Toledo Palomino, Ekaterina Durmanova, Seiji Isotani, and Lennart E. Nacke. 2024. An Umbrella Review of Reporting Quality in CHI Systematic Reviews: Guiding Questions and Best Practices for HCI. *ACM Trans. Comput.-Hum. Interact.* 31, 5, Article 57 (Nov. 2024), 55 pages. doi:10.1145/3685266

[212] Stephen Rollnick, William R Miller, Christopher C Butler, and Mark S Aloia. 2008. Motivational interviewing in health care: helping patients change behavior.

[213] * John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1163–1172.

[214] * John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2015. Pass the ball: enforced turn-taking in activity tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2417–2426.

[215] * Kathleen Ryan, Samantha Dockray, and Conor Linehan. 2022. Understanding how ehealth coaches tailor support for weight loss: towards the design of person-centered coaching systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[216] * Herman Saksono, Carmen Castaneda-Sceppa, Jessica Hoffman, Vivien Morris, Magy Seif El-Nasr, and Andrea G Parker. 2020. Storywell: designing for family fitness app motivation by using social rewards and reflection. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[217] * Herman Saksono, Carmen Castaneda-Sceppa, Jessica Hoffman, Magy Seif El-Nasr, Vivien Morris, and Andrea G Parker. 2019. Social reflections on fitness tracking data: A study with families in low-SES neighborhoods. In *Proceedings of the 2019 chi conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[218] * Herman Saksono, Carmen Castaneda-Sceppa, Jessica A Hoffman, Magy Seif El-Nasr, and Andrea Parker. 2021. StoryMap: Using Social Modeling and Self-Modeling to Support Physical Activity Among Families of Low-SES Backgrounds. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[219] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja SB Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. 2023. Changes in research ethics, openness, and transparency in empirical studies between CHI 2017 and CHI 2022. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–23.

[220] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, et al. 2019. HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[221] * Jomara Sandbulte, Chun-Hua Tsai, and John M Carroll. 2021. Working Together in a PhamilySpace: Facilitating Collaboration on Healthy Behaviors Over Distance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–32.

[222] Gavin Sandercock. 2024. The Standard Error/Standard Deviation Mix-Up: Potential Impacts on Meta-Analyses in Sports Medicine. *Sports Medicine* 54 (2024), 1–10.

[223] * Kim Sauvé, Saskia Bakker, and Steven Houben. 2020. Econundrum: Visualizing the climate impact of dietary choice through a shared data sculpture. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 1287–1300.

[224] ** Hanna Schäfer and Martijn C Willemsen. 2019. Rasch-based tailored goals for nutrition assistance systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 18–29.

[225] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 5412–5427.

[226] * Zachary Schmitt and Svetlana Yarosh. 2018. Participatory design of technologies to support recovery from substance use disorders. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.

[227] * Jessica Schroeder, Chia-Fang Chung, Daniel A Epstein, Ravi Karkar, Adele Parsons, Natalia Murinova, James Fogarty, and Sean A Munson. 2018. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. In *Proceedings of the 2018 designing interactive systems conference*. Association for Computing Machinery, New York, NY, USA, 135–148.

[228] * Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A Munson. 2019. Examining opportunities for goal-directed self-tracking to support chronic condition management. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 4 (2019), 1–26.

[229] * Jessica Schroeder, Chelsey Wilkes, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M Linehan. 2018. Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[230] Taylor Jackson Scott, Katie Kuksenok, Daniel Perry, Michael Brooks, Ona Anicello, and Cecilia Aragon. 2012. Adapting grounded theory to construct a taxonomy of affect in collaborative online chat. In *Proceedings of the 30th ACM international conference on Design of communication*. Association for Computing Machinery, New York, NY, USA, 197–204.

[231] * Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong'Cherie' Chen, Likang Sun, and Geoff Kaufman. 2019. Designing user interface elements to improve the quality and civility of discourse in online commenting behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[232] Sneha Shankar, Sheila K Marshall, and Bruno D Zumbo. 2020. A systematic review of validation practices for the goal attainment scaling measure. *Journal of Psychoeducational Assessment* 38, 2 (2020), 236–255.

[233] Garriy Shteynberg and Adam D Galinsky. 2011. Implicit coordination: Sharing goals with similar others intensifies goal pursuit. *Journal of Experimental Social Psychology* 47, 6 (2011), 1291–1294.

[234] * Aneesha Singh, Annina Klapper, Jinni Jia, Antonio Fidalgo, Ana Tajadura-Jiménez, Natalie Kanakam, Nadia Bianchi-Berthouze, and Amanda Williams. 2014. Motivating people with chronic pain to do physical activity: opportunities for technology design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2803–2812.

[235] * Manya Sleeper, Alessandro Acquisti, Lorrie Faith Cranor, Patrick Gage Kelley, Sean A Munson, and Norman Sadeh. 2015. I Would Like To..., I Shouldn't..., I Wish I... Exploring Behavior-Change Goals for Social Networking Sites. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. Association for Computing Machinery, New York, NY, USA, 1058–1069.

[236] Petr Slovak and Sean A Munson. 2024. HCI Contributions in Mental Health: A Modular Framework to Guide Psychosocial Intervention Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–21.

[237] Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (2019), 62–65.

[238] ** Katarzyna Stawarz, Anna L Cox, and Ann Blandford. 2015. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2653–2662.

[239] * Evropi Stefanidi, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2023. Children with ADHD and their care ecosystem: Designing beyond symptoms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[240] Evropi Stefanidi, Marit Bentvelzen, Paweł W. Woźniak, Thomas Kosch, Mikołaj P. Woźniak, Thomas Mildner, Stefan Schneegass, Heiko Müller, and Jasmin Niess. 2023. Literature Reviews in HCI: A Review of Reviews. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 509, 24 pages. doi:10.1145/3544548.3581332

[241] * Elizabeth Stowell, Yixuan Zhang, Carmen Castaneda-Sceppa, Margie Lachman, and Andrea G Parker. 2019. Caring for Alzheimer's disease caregivers: A qualitative study investigating opportunities for Exergame innovation. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.

[242] * Lie Ming Tang and Judy Kay. 2017. Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 2 (2017), 1–28.

[243] * Lie Ming Tang and Judy Kay. 2018. Scaffolding for an olm for long-term physical activity goals. In *Proceedings of the 26th conference on user modeling, adaptation and personalization*. Association for Computing Machinery, New York, NY, USA, 147–156.

[244] Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Trans. Comput.-Hum. Interact.* 27, 5, Article 34 (Aug. 2020), 53 pages. doi:10.1145/3398069

[245] * Helma Torkamaan and Jürgen Ziegler. 2022. Recommendations as challenges: estimating required effort and user ability for health behavior change recommendations. In *27th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 106–119.

[246] III Turner, Herbert M and Robert M Bernard. 2006. Calculating and synthesizing effect sizes. *Contemporary issues in communication science and disorders* 33, Spring (2006), 42–55.

[247] Muhammad Umair, Niaz Chalabianloo, Corina Sas, and Cem Ersoy. 2021. HRV and stress: a mixed-methods approach for comparison of wearable heart rate sensors for biofeedback. *IEEE Access* 9 (2021), 14005–14024.

[248] Muhammad Umair, Corina Sas, Niaz Chalabianloo, and Cem Ersoy. 2021. Exploring personalized vibrotactile and thermal patterns for affect regulation. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 891–906.

[249] United States Census Bureau. 2023. *Census United States Older Population*. United States Census Bureau. https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html

[250] U.S. Forum on Child and Family Statistics. 2024. *America's Children*. U.S. Forum on Child and Family Statistics. https://www.childstats.gov/americaschildren/demo.asp

[251] * Stephen Uzor and Lynne Baillie. 2018. Exploring the communication of progress in home-based falls rehabilitation using exergame technologies. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.

[252] * Robby van Delden, Danny Plass-Oude Bos, Antje Jacoba Vivian de With, Koen Vogel, Randy Klaassen, Nynke Zwart, Joyce Faber, Boony Thio, and Mattienne van der Kamp. 2020. SpiroPlay, a suite of breathing games for spirometry by kids & experts. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, New York, NY, USA, 400–413.

[253] ** Loes Van Renswouw, Steven Vos, Pieter Van Wesemael, and Carine Lallemand. 2021. Exploring the Design Space of InterActive Urban Environments: Triggering physical activity through embedded technology. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 955–969.

[254] ** Madhurima Vardhan, Narayan Hegde, Srujana Merugu, Shantanu Prabhat, Deepak Nathani, Martin Seneviratne, Nur Muhammad, Pranay Reddy, Sriram Lakshminarasimhan, Rahul Singh, et al. 2022. Walking with pace-personalized and automated coaching engine. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 57–68.

[255] * Gabriela Villalobos-Zúñiga, Iyubanit Rodríguez, Anton Fedosov, and Mauro Cherubini. 2021. Informed Choices, Progress Monitoring and Comparison with Peers: Features to Support the Autonomy, Competence and Relatedness Needs, as Suggested by the Self-Determination Theory. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. Association for Computing Machinery, New York, NY, USA, 1–14.

[256] * Thiemo Wambsganss, Andreas Janson, Tanja Käser, and Jan Marco Leimeister. 2022. Improving students argumentation learning with adaptive self-evaluation nudging. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[257] * Dennis Wang, Marawin Chheang, Siyun Ji, Ryan Mohta, and Daniel A Epstein. 2022. SnapPI: Understanding Everyday Use of Personal Informatics Data Stickers on Ephemeral Social Media. *Proceedings of the ACM on human-computer interaction* 6, CSCW2 (2022), 1–27.

[258] * Kendra A Wannamaker, Sandeep Zechariah George Kollannur, Marian Dörk, and Wesley Willett. 2021. I/o bits: User-driven, situated, and dedicated self-tracking. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 523–537.

[259] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44.

[260] World Health Organization. 2024. *Ageing and Health*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[261] * Alan Yusheng Wu and Cosmin Munteanu. 2018. Understanding older users' acceptance of wearable interfaces for sensor-based fall risk assessment. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[262] * Kefan Xu, Xinghui Yan, and Mark W Newman. 2022. Understanding people's experience for physical activity planning and exploring the impact of historical records on plan creation and execution. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–15.

[263] ** Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: leveraging just-in-time self-affirmation for smartphone overuse reduction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[264] * Renwen Zhang, Kathryn E. Ringland, Melina Paan, David C. Mohr, and Madhu Reddy. 2021. Designing for emotional well-being: integrating persuasion and customization into mental health technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13.

[265] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *Proceedings of the 2022 CHI Conference on Human*

Factors in Computing Systems (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 570, 24 pages. doi:10.1145/3491102.3501855

[266] * Shuo Zhou and Timothy Bickmore. 2022. A virtual counselor for breast cancer genetic counseling: adaptive pedagogy leads to greater knowledge gain. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.

[267] * Xingchen Zhou, Pei-Luen Patrick Rau, and Xueqian Liu. 2021. " Time to Take a Break" How Heavy Adult Gamers React to a Built-In Gaming Gradual Intervention System. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–30.