Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis

(Article begins on next page)

Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis

ALBERTO MONGE ROFFARELLO, Politecnico di Torino, Italy
LUIGI DE RUSSIS, Politecnico di Torino, Italy

Public media and researchers in different areas have recently focused on perhaps unexpected problems that derive from an excessive and frequent use of technology, giving rise to a new kind of psychological "digital" wellbeing. Such a novel and pressing topic has fostered, both in the academia and in the industry, the emergence of a variety of *digital self-control tools* allowing users to self-regulate their technology use through interventions like timers and lock-out mechanisms. While these emerging technologies for behavior change hold great promise to support people's digital wellbeing, we still have a limited understanding of their real effectiveness, as well as of how to best design and evaluate them. Aiming to guide future research in this important domain, this article presents a systematic review and a meta-analysis of current work on tools for digital self-control. We surface motivations, strategies, design choices, and challenges that characterize the design, development, and evaluation of digital self-control tools. Furthermore, we estimate their overall effect size on reducing (unwanted) technology use through a meta-analysis. By discussing our findings, we provide insights on how to (i) overcome a limited perspective that exclusively focuses on technology overuse and self-monitoring tools, (ii) evaluate digital self-control tools through long-term studies and standardized measures, and (iii) bring ethics in the digital wellbeing discourse and deal with the business model of contemporary tech companies.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**.

Additional Key Words and Phrases: digital self-control tools, digital wellbeing, digital overuse, behavior change, persuasive technology

ACM Reference Format:
Alberto Monge Roffarello and Luigi De Russis. 2020. Achieving Digital Wellbeing Through Digital Self-Control Tools: A Systematic Review and Meta-Analysis. *ACM Trans. Comput.-Hum. Interact.* 00, 0, Article 000 (2020), 66 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

A growing — yet debated — discussion around the negative aspects of overusing technology is nowadays led by both mainstream media [4, 42] and researchers working in different areas, from addictive behaviors and disorders [17, 119] to psychology [151] and HCI [25, 228]. Despite a large part of this narrative is often speculative [215], especially when everyday behaviors like smartphone use are associated to an addiction framing [123, 124], there is an established body of evidence suggesting that an excessive and frequent use of technological sources like mobile devices [15, 122], social media [148], and the Internet in general [226] may negatively interfere with daily activities and ongoing tasks such as studying [15], driving [58], and even sleeping [122], to the point of

Authors' addresses: Alberto Monge Roffarello, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, TO, 10129, Italy, alberto.monge@polito.it; Luigi De Russis, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, TO, 10129, Italy, luigi.derussis@polito.it.

1073-0516/2020/0-ART000 $15.00
https://doi.org/10.1145/1122445.1122456

creating problems for people's mental health [122] and social interactions [66, 216]. It is nowadays clear that many people feel conflicted about the amount of time they spend on their devices [129], especially when technological sources are used passively [91, 218]. This body of evidence has recently prompted researchers to consider a new kind of psychological wellbeing affecting today's individuals, the so-called *digital wellbeing*. Burr et al. [39] defines it as "the impact of digital technologies on what it means to live a life that is good for a human being in an information society." As conceptualized by Cecchinato et al. [41], digital wellbeing can be framed from multiple points of view, including medical-oriented, user-oriented, and design-oriented perspectives [41], and even tech giants like Google [5] and Apple [6] have recently embraced, at least in theory, a new design philosophy "to giving everyone the tools they need to develop their own sense of digital wellbeing [5]."

With the aim of promoting people's digital wellbeing, the last few years have seen the flourishing of *Digital Self-Control Tools (DSCTs)* both in academia [103, 105] and as off-the-shelf products [87, 201], with some commercial apps, e.g., Forest [201], that have gathered millions of users [143]. DSCTs aim at supporting self-control over technology use by allowing users to track their usage patterns and to define interventions, e.g., timers and lock-out mechanisms [159], on their different devices and online services. As they are ideally intended to help users improve their behaviors with technology, DSCTs can be viewed, from a broader perspective, as a technology for behavior change and sustainment [77], or, more specifically, as digital interventions for behavior change [178].

Despite the potential opportunities for designing technology that promotes digital wellbeing for their users, this is an emerging research area, and the development of effective applications in this field is tied up to several complex and intertwined challenges. While a plethora of DSCTs have been developed over the past years, in particular, we still have a limited understanding of how to best design and evaluate them [49, 149], and the real effectiveness of these tools, especially in the long-term, is yet underexplored [116, 117]. The digital wellbeing context, indeed, is characterized by specific peculiarities that differentiate it from other behavior change domains like eating and physical activities, e.g., the fact that technological sources are often, at the same time, the source of the problem and the platform with which the interventions are delivered to the user [198]. Despite the prevalence of research works investigating problems like technology overuse, and some insightful recent reviews [144, 159] on the strategies adopted (and not adopted) by commercially available DSCTs, a comprehensive overview of all the phases that characterize the design process of DSCTs, from the initial idea to their evaluation, is currently missing. Furthermore, as we consider the development of these emerging technologies a delicate process that may involve vulnerable user groups, e.g., minors, and their sensitive information, e.g., device usage data, there is also the urgent need to understand to what extent ethical considerations and issues are taken into account in the digital wellbeing context.

To assess the state-of-the-art characterizing the development of DSCTs and to guide future research in the important and pressing topic of digital wellbeing, this paper reports on the results of a systematic review from the computing and HCI literature of current work discussing and/or proposing tools for digital self-control. Table 1 reports the research questions that guide our analysis.

We first analyze the *digital wellbeing aspects* and the *research goals* that motivate researchers to discuss, design, and evaluate DSCTs (**RQ1**), and we extract the *strategies* that researchers are proposing to design interventions for digital self-control (**RQ2**). At the same time, we also assess whether the proposed interventions and their evaluations are theoretically grounded, e.g., with designs that are informed by behavioral theories and/or specific constructs. Then, we report on the *challenges* that characterize the field of tools for digital self-control, and we describe the extent to which the analyzed papers take into account *ethical* issues and/or discuss ethical implications (**RQ3**). We conclude our review by analyzing how researchers *evaluate* their DSCTs (**RQ4**). We

Table 1. The research questions investigated in our work. We explored the field of interventions for digital self-control by analyzing underlying *motivations*, adopted *strategies*, experienced *challenges*, *ethical* issues and implications, and reported *evaluations*.

#	Category	Research Question	Section
RQ1	Motivations	Which research goals guide the discussion and/or the proposal of digital self-control tools and what aspects of digital wellbeing are researchers in this field targeting?	Section 3
RQ2	Strategies	Which interventions for digital self-control are researcher discussing and/or proposing, and which behavioral theories and constructs, if any, are researcher considering?	Section 4
RQ3	Challenges & Ethics	Which challenges characterize the field of tools for digital self-control, and to what extent do researchers describe ethical challenges or implications when discussing or proposing these tools?	Section 5
RQ4	Evaluation	How do researcher evaluate digital self-control tools? Are digital self-control tools effective in reducing people's time using devices and/or applications?	Section 6

extract, in particular, information about participants and recruiting processes, collected measures, and study designs. We also report on the results of a meta-analysis with which we quantitatively assess the effectiveness of DSCTs in reducing the time spent by users on target devices, websites, and/or mobile applications.

Overall, our work shows that DSCTs target different aspects of people's digital wellbeing, from reducing digital overuse, distractions, and dark patterns to improving the quality of the interaction, e.g., in terms of meaningfulness. Our meta-analysis, in particular, demonstrates that DSCTs have a small to medium effect on reducing the time spent by users on distractive technological sources. Driven by different underlying research goals, from designing novel interventions to understanding how users respond to the usage of existing DSCTs, researchers proposed and explored a variety of intervention strategies to support and continuously improve this promising result, although with different approaches. Papers describing novel DSCT implementations, indeed, mainly focus on block/removal strategies, e.g., timers and lockout mechanisms. This confirms previous work on commercially available DSCTs [144, 159], and, together with the challenges that the same researchers highlight in their papers, opens the way to investigate more intelligent tools that can learn from the user and adapt to different devices, usage patterns, and people's physical capabilities and technology competences. On the contrary, the papers in our corpus that include more generic studies, i.e., without any particular DSCT implementation, explicitly call researchers to explore alternative approaches to blocking interactions, e.g., designing for meaningfulness.

Our work also highlights three main important gaps in the current literature on DSCTs. First, as in other behavior change domains [155, 156], the research on digital self-control tools suffers from a "theoretical gap [77]," with the majority of the analyzed papers that do not mention any behavioral theory, and other works that do not provide sufficient evidence on how the adopted theories relate to the proposed interventions. Another gap is related to the evaluation of DSCTs, which is nearly always based on short-term experiments that lack important methodological aspects like control groups and follow-up assessments, therefore casting doubts on the promising effects reported in the papers under analysis. Echoing recent reviews in similar domains (e.g., [191, 210]), the last gap highlights a limited interest in reporting ethical considerations in papers describing DSCTs. Stemming from the reported findings, our final discussion points to (i) the opportunity of overcoming a limited perspective that exclusively focuses on technology overuse and self-monitoring tools, (ii) the challenge of improving the evaluation of DSCTs through longer-term studies and standardized measures, and (iii) the need for bringing ethics in the digital wellbeing discourse. For this last point in particular, we call the HCI community and technology companies to work together to resolve the inherent contradiction of designing for digital wellbeing in a business model which currently incentives frequent and continuous usage.

To our knowledge, this is the first systematic literature review of research works on digital self-control, and the first attempt to estimate the overall effect size of DSCTs. While some of the findings included in this paper may have been discussed in previous works, our work rigorously and systematically evidences key gaps and issues in the digital wellbeing literature that otherwise would have been difficult to reconcile. The paper is organized as follow. Section 2 describes the methodologies used for our review. Section 3 reports on the digital wellbeing aspects and underlying research goals (**RQ1**). Section 4 focuses on adopted strategies and applied theories (**RQ2**). Section 5 highlights challenges and ethical implications (**RQ3**). Section 6 focuses on evaluations (**RQ4**). Section 7 discusses the implications of our work and possible future directions for the digital wellbeing research area. Eventually, Section 8 concludes the paper.

2 METHODOLOGY

To identify and select relevant papers for our systematic literature review, we followed the PRISMA literature review guidelines [130, 158] (see Figure 1 for our procedural flowchart).

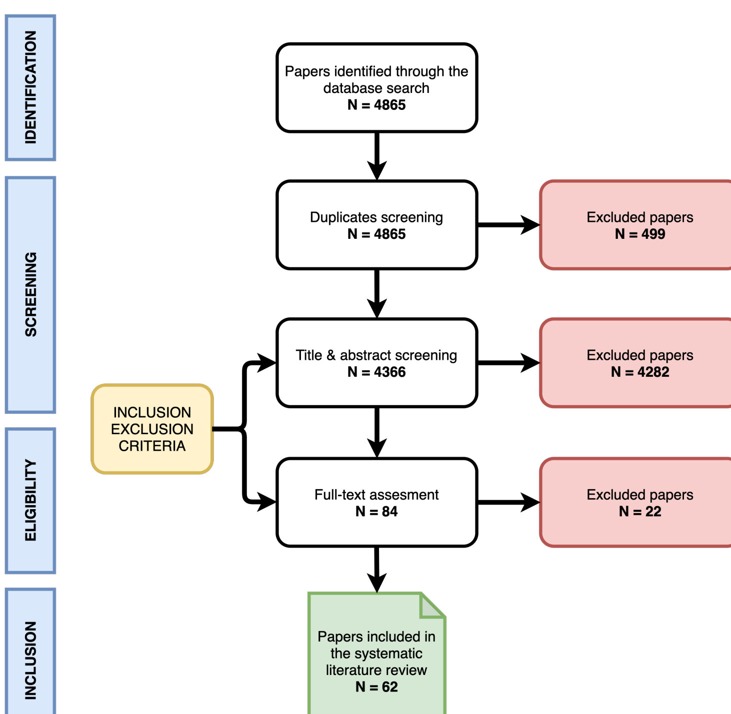

Fig. 1. Procedural flowchart following the PRISMA guidelines.

Table 2. The search queries used to search the electronic database of the ACM Guide to the Computing Literature. We constrained the all the searches by specifying a custom time range for publication (January 2000 to June 2021) and a specific "content type" for the manuscripts ("Research Article").

Search Query	# Results
"behavior change" OR "self control" OR "self regulation"	4011
"internet addiction" OR "smartphone addiction" OR "social media addiction" OR "technology addiction" OR "app addiction"	443
"digital intervention*" OR "digital nudge*"	182
"digital break*" OR "digital diet" OR "behavior restriction*"	49
"smartphone overload" OR "smartphone overuse" OR "phone overload" OR "phone overuse"	49
"digital wellbeing"	42
"digital overload" OR "digital overuse" OR "technology overload" OR "technology overuse" OR "digital distraction*"	41
"digital distraction*"	22
"internet overload" OR "internet overuse"	21
"social media overload" OR "social media overuse" OR "social networks overload" OR "social networks overuse"	5

2.1 Paper Selection

As in recent literature reviews in the HCI domain (e.g., [210]), we identified relevant papers by searching the electronic database of the Association for Computing Machinery (ACM) Guide to the Computing Literature[1]. According to its specifications, such a database is the most comprehensive bibliographic source collection in the field of computing and HCI research. It integrates the traditional ACM Digital Library with conference proceedings, journals, magazines, books, and abstracts of key publishers like IEEE, Springer, and Elsevier. We explicitly focused on research works, only, as our aim was to investigate a broad set of research questions specifically targeted to the digital wellbeing research area, e.g., research challenges and ethical considerations. As reported in the following sections, however, our review also includes commercially available tools that have been evaluated and/or discussed in the mentioned research studies.

The final corpus of papers presented in this article is the result of a search conducted on the 18th of June 2021. We used the work of Lyngs et al. [144] as a reference point to build our search. Such a paper bases its analysis on "17 HCI papers which have either built novel design intervention or evaluated existing interventions to support self-control over digital device use" (p. 2). We defined our search terms (from "behavior change" to "social network overuse", Table 2) looking at words in titles, abstracts, and keywords in the 17 papers analyzed by Lyngs et al. [144], and we further enriched them by using subject terms, e.g., to capture "Internet overload" as well as "Internet overuse." Furthermore, we also built a validation set for our search by including the same 17 papers. Overall, the initial search identified a total of 4,865 records. The extracted records included all the papers of the validation set, thus confirming the completeness of our search terms. One of the authors analyzed the retrieved collection by removing 499 duplicates. Other 4,282 records were excluded through a screening of the titles and the abstracts. Such a screening was performed to determine whether the collected papers were related to the the fields of digital wellbeing and digital self-control. Papers were eligible for inclusion if they targeted any aspects related to the digital wellbeing, e.g., overusing technology or digital distractions, and if they discussed or analyzed, even just qualitatively, at least an intervention for behavior change. Papers were instead excluded if they fell into one of these exclusion criteria:

[1]https://libraries.acm.org/digital-library/acm-guide-to-computing-literature, last visited on July 7, 2021.

- papers discussing the topic of digital wellbeing in a generic way, without focusing on any behavior change interventions;
- papers discussing or analyzing behavior change interventions that do not explicitly target the usage of devices and/or online services, e.g., interventions against sedentary lifestyles;
- papers presenting pure parental-control solutions, with interventions that are not defined by the users themselves but rather externally imposed, e.g., by parents;
- limited research reports, magazines, and editorials.

At the end of the screening, we further analyzed whether the remaining 82 papers respected the defined inclusion and exclusion criteria. Through a final full-text assessment, in particular, we removed another set of papers. The final set of 62 articles[2] included in the systematic literature review (*General Corpus*) was composed of 51 conference papers (38 full papers, 7 posters, and 6 extended abstracts) and 11 journal papers. We checked all the conference and journal websites to ensure that all the considered papers, including posters and extended abstracts, had been rigorously peer-reviewed. The most common conferences were CHI[3] (29), UbiComp[4] (6), and CSCW[5] (2), while journal articles mostly came from IMWUT[6] (5). As reported in Figure 2, our final corpus shows that the filed of tools for digital self-control is a relatively new research area characterizing the last 10 years, only.

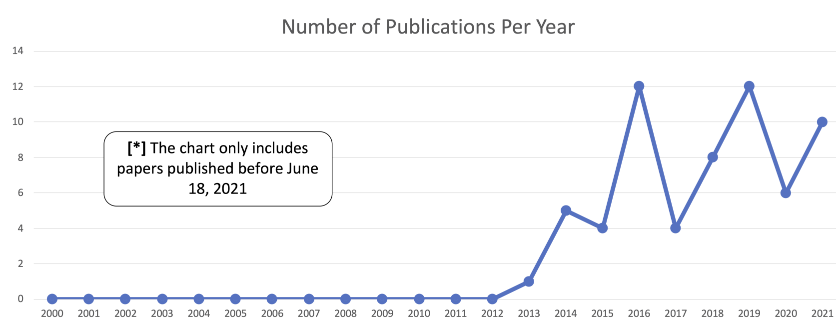

Fig. 2. Number of publications that fulfill our inclusion/exclusion criteria per year. The graph highlights that the filed of tools for digital self-control is a relatively new research area.

To conduct additional and specific analysis, we also identified 2 main subsets of contributions in our *General Corpus*:

Qualitative Discussions (N=17): papers that analyze some aspects related to the digital wellbeing context, e.g., the usage of a given device [162] or the motivations for (not) using a social network [195], to inform a qualitative discussion around possible interventions for digital self-control that could be implemented in the future.

Implemented Tools (N=45): papers that focus on the implementation of a DSCT. This includes papers presenting the implementation of novel interventions in a dedicated tool (e.g., [113]) and papers analyzing the usage of existing intervention strategies (e.g., [159]) or commercially available DSCTs (e.g., [47]).

[2] The included papers are highlighted in the References list through a check mark (✓).
[3] https://dl.acm.org/conference/chi, last visited on July 8, 2021
[4] https://dl.acm.org/conference/UbiComp, last visited on July 8, 2021
[5] https://dl.acm.org/conference/CSCW, last visited on July 8, 2021
[6] https://dl.acm.org/journal/imwut, last visited on July 8, 2021. The IMWUT journal publishes the full papers of the UbiComp conference since 2017.

In addition, we extracted a further subset of papers from the *Implemented Tools* corpus:

Evaluated Tools (N=37): a subset of the Implemented Tools corpus that includes papers reporting on at least an evaluation of the proposed or analyzed DSCT implementation, e.g., an in-the-wild [149] or an in-the-lab study [161].

2.2 Data Extraction & Coding Process

To systematically extract data from our corpus, we created a data extraction sheet by coding different aspects of our research questions. We used the first columns of the sheet to characterize the papers under analysis by their authors, title, abstract, publication type and year, and to summarize the presented tools. We extracted the *motivations* guiding the discussion and the analysis of DSCTs, including the targeted digital wellbeing aspects (**RQ1**), and we recorded information about the *strategies* investigated by researchers in their works, including the applied theories, if any (**RQ2**). In analyzing the strategies, in particular, we distinguished between papers that included an implementation of a DSCT (*Implemented Tools*) and papers discussing interventions qualitatively, only (*Qualitative Discussions*). For the former, we extracted tool-related information like the delivery platform and the followed design process. For the latter, instead, we coded the proposed interventions to understand whether there are promising strategies that have not yet been implemented and tested. For each paper included in our *General Corpus*, we also noted the *challenges* encountered by researchers in studying digital self-control interventions to achieve digital wellbeing, as well as the described ethical issues or implications, if discussed (**RQ3**). Regarding the *evaluation* of DSCTs (**RQ4**), we focused on the *Evaluated Tools* corpus, by extracting information about the conducted studies. This included general information about the implemented studies (i.e., number, duration, number of participants and their demographics), as well as information about how the studies were conducted (i.e., recruiting processes, collected metrics, study design, usage of control groups and follow-up sessions). Furthermore, we extracted all the quantitative results reported in the analyzed papers to run the meta-analysis on DSCTs effectiveness (see Section 6.2).

Besides the described aspects, we also applied two state-of-the-art coding schemes:

- To further investigate the characteristics of the *Implemented Tools* (**RQ2**), we applied a coding scheme resulting from a recent review of commercially available DSCTs [143]. According to the scheme, *block/removal* features aim to assist users in avoiding distractions, e.g., through blocking distracting functionality or by allowing users to set up limits on how much time can be spent on a given website or mobile application. *Self-tracking* means tracking user's behavior and providing feedback, e.g., through visualisations of the captured data, timers, and countdowns. *Goal advancement* allows users to explicitly specify an objective and to follow its advancement, e.g., through a reminder of a concrete time goal. *Reward/punishment* refers to providing some rewards or punishments for the way in which a device or one of its specific services, e.g., a mobile app, is used.
- To further investigate ethical principles discussed in our *General Corpus* (**RQ4**), we applied a framework of four broad moral principles informed by healthcare ethics [27] and already adopted in other HCI literature reviews, e.g., [191]. According to the framework, *autonomy* refers to the respect for people's decision-making ability. This is achieved by supporting the understanding of information, e.g., through informed consent. *Non-maleficence* is the explicit intention of not causing harm through the designed intervention. *Beneficence* means trying to provide benefits and balancing benefits against risks and costs with the aim of preventing harm. *Justice* refers to a fair distribution of the benefits, risks, and costs of an intervention across different populations, by avoiding any discrimination based on users' characteristics like social class, race, and gender.

Table 3. The digital wellbeing aspects targeted by the research works on digital self-control included in our *General Corpus*.

Digital Wellbeing Aspect	Description
Digital Overuse (N=34)	The abundance of digital information and communication options, and pressure to use them effectively and constantly.
Distractions & Productivity (N=16)	External and self-interruptions (e.g., notifications and checking emails, respectively), and how these distractions influence users' productivity.
Digital Use (N=8)	How devices, online services, and/or DSCTs are used by people, and how these usages could be improved to positively influence users' digital wellbeing.
Attention-Capture Dark Patterns (N=2)	Designs and functionality that make users do things that they did not mean to, with the final aim of maximizing the time spent by the same users on mobile apps and websites.
Environmental Footprints (N=2)	The impact of Internet and its infrastructure on the environment, e.g., in terms of global greenhouse gas emissions.
Dangerous Habits (N=1)	Behaviors with digital devices in a given contextual situation that pose a risk to personal safety, e.g., using the smartphone while driving.

The described extraction sheet template was created by one of the authors by coding ten randomly selected papers. The sheet was then checked by the second author, who implemented some minor adjustments. Each paper of the corpus was finally analyzed using the final version of the extraction sheet template.

3 MOTIVATIONS FOR HAVING DIGITAL-SELF CONTROL TOOLS

This section describes what aspects of digital wellbeing are targeted by the research studies on digital self-control included in our *General Corpus*, as well as the underlying research goals guiding the discussion and/or the design of the proposed interventions (**RQ1**).

3.1 Targeted Digital Wellbeing Aspects

Table 3 summarizes the digital wellbeing aspects targeted by the research works on digital self-control included in our *General Corpus*.

The majority of the analyzed papers ($N = 34$, [16, 19, 43, 62, 64, 81, 82, 105, 106, 112, 114, 116–118, 127, 132, 134, 144, 145, 159, 166, 167, 171–173, 177, 188, 189, 195, 198, 202, 205, 213, 229]) focus on *digital overuse*, i.e., "the abundance of digital information and communication options, and pressure to use them effectively and constantly [36]." While users can benefit from the use of digital technologies, e.g., through new opportunities for social support [211], an overuse of Internet-enabled devices and services may however negatively impact subjective wellbeing [71]. Since the smartphone is often considered as the major source of digital wellbeing problems [159], e.g., due to its portable and interactive characteristics [129], the majority of papers targeting digital overuse ($N = 26$) propose digital interventions to assist users in controlling the usage of this kind of device [16, 19, 62, 64, 105, 112, 114, 127, 144, 159, 171, 172, 177, 213] and/or of specific mobile applications [81, 82, 106, 115, 116, 134, 139, 159, 166, 167, 188, 202], with the study of Lyngs et al. [139] that specifically investigates how to redesign the mobile app of YouTube to support users' sense of agency. Some of the DSCT implementations proposed in these papers, e.g., *HabitLab* [116] and *ScreenLife* [189], also include a browser extension or an application at the operating system level to intervene on both smartphones and PCs. Researchers in the digital wellbeing context started to explore such a multi-device perspective only recently (see the study of Monge Roffarello and De Russis [160] for instance). As users typically use more than one device at a time [3], however, DSCTs targeting more than one device are fundamental to capture all the nuances of people's digital wellbeing [125], and, at the same time, to avoid "deceptive" behaviors like using the smartphone to overcome a block on the PC [186]. The remaining 12 papers targeting digital overuse focus on the PC, only, by allowing users to mitigate the overuse of websites [43, 117, 118, 144, 189, 198], online gaming [173], and specific social networks like Facebook [116, 145], Twitter [195, 205], and WeChat [132].

Interestingly, instead of focusing on *over*-use, 8 papers in our *General Corpus* [65, 83, 140, 160, 162, 165, 189, 204] propose interventions for digital self-control by focusing on *use*, i.e., by analyzing how devices, online services, and/or DSCTs are used, and how these usages could be improved to positively influence users' digital wellbeing. In particular, the work by Lukoff et al. [140] distinguishes between *habitual* use and *instrumental* use. Habitual use is often associated to a meaningless experience, e.g., browsing social media to pass the time. Instrumental use, instead, often results in meaningful experiences, e.g., using an app to achieve a specific and well-defined goal. Furthermore, papers focusing on use are not restricted to single users, only, but they also explore how technology use influence social situations. As an example, Hiniker et al. [83] discuss possible solutions to improve mobile phone use by exploring how adults use their smartphones while caring for children at the playground. Moser et al. [162], instead, base their discussion on interventions for digital self-control by investigating the users' attitudes towards the usage of smartphones at mealtimes.

Another large set of papers ($N = 16$, [11, 33, 47, 88, 101–104, 108, 113, 133, 149, 150, 186, 214, 223]) describe solutions to assist users in avoiding *distractions*, with the aim of improving users' *productivity*. Digital wellbeing is, indeed, strictly related to productivity [41], with self-interruptions and external notifications that may impact productive activities like studying and working [160]. Typically, implementations of DSCTs targeting distractions, e.g., *Knob* [186] and *UpTime* [214], are designed to help office workers, therefore it is not surprising that the majority of them focus on blocking distractive websites on the computer's browser [11, 47, 104, 108, 133, 149, 186, 214, 223]. Instead of blocking distractive sources, *TimeToFocus* [33] visualizes, in real time, a browser notification that shows how long the user switches away from a main task window, while the tool presented in [88], i.e., *Aiki*, redirects users to a productive website before letting them to access a distractive website. Other papers [101–103, 113], instead, describe and propose tools and solutions that aim at minimizing smartphone distractions, e.g., hrough the block of specific mobile apps and/or notifications [103]. Some of the aforementioned papers also present DSCTs that are able to limit distractions coming both from smartphones and PCs [47, 104, 186]. An example is *RecueTime* [185] (described in [47]), a commercial DSCT with which users can block distractive websites on the PC while putting the smartphone in do-not-disturb mode.

With a more focused view on aspects like overuse and distractions, Kollnig et al. [115] and Lukoff et al. [139] developed and proposed, respectively, digital self-control interventions to alleviate *attention-capture dark patterns* [139]. Dark patterns have been originally defined by Brignull [34] as "tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something." Such malicious patterns take advantage of cognitive biases of end users and can have detrimental effects on people's lives [115]. Clearly, dark patters can be specialized in the digital wellbeing context, too. The work of Kollnig et al. [115], for example, explores how dark patterns encouraging more user interaction on social network mobile apps, e.g., stories and notification counters, can be removed. With a similar goal but specifically targeted to the YouTube

Table 4. The themes extracted in our analysis on the research goals of papers related to the filed of digital self-control.

General Themes	Specific Themes
Designing Novel Interventions (N=29)	New interventions strategies (N=11) Users' contexts and activities (N=6) Generalizable solutions (N=6)
Understanding Interventions (N=14)	Existing tools (N= 7) Reaction to novel strategies (N=8)
Understanding Users & Patterns (N=19)	Technology use analysis (N=11) DSCTs use analysis (N=8)

mobile app, Lukoff et al. [139] defined attention-capture dark patterns as "designs that manipulate the user into spending time and attention in an app against their best interests" (p. 2).

Another aspect covered by a small but interesting set of papers in the *Qualitative Discussions* corpus ($N = 2$, [80, 224]) is the need of developing DSCTs that effectively allow users to reduce tech-usage at a global scale, with the aim of limiting the *environmental footprints* of the infrastructure made of Internet and digital devices. As highlighted by Widdicks and Pargman [224], indeed, such an infrastructure is growing at an increasing and unsustainable rate by following a "Cornucopian paradigm." Taking inspiration from the work of Preist et al. [181], Hill et al. [80] explains such a paradigm in the introduction of their work: "efficiencies in the infrastructure allow for new data-intensive services to be designed, which drives demand for such services, and ultimately leads to further growth and required efficiencies in the infrastructure" (p. 204).

Finally, the remaining paper in our *General Corpus* (Morley et al., [161]) focuses on a specific aspect of the digital wellbeing. The authors, in particular, developed an immersive user experience based on virtual reality to teach users how to avoid *dangerous habits* like using the smartphone while driving. Behaviors like writing a message while driving, indeed, are often performed without awareness [26], and they can have serious consequences on people's safety.

3.2 Research Goals of Studies on Digital Self-Control

Table 4 summarizes the themes that describe the underlying goals of the research studies included in our *General Corpus*. We coded goals in terms of *designing novel interventions*, *understanding interventions*, and *understanding users & patterns*.

3.2.1 Designing Novel Interventions. A common goal that is shared across a large proportion of the papers in our *General Corpus* ($N = 29$, [11, 64, 80, 82, 88, 102, 103, 105, 106, 113–115, 127, 134, 139, 160, 161, 166, 167, 171–173, 177, 186, 198, 202, 205, 214, 224]) is the need of designing and evaluating *novel interventions* that improve people's digital wellbeing in different contexts. Since digital wellbeing is a relatively recent topic, researchers are continuously investigating new strategies to assist users in self-regulating the usage of their devices and services, with the aim of improving user's self-discipline and self-improvement [202]. Thirteen papers that propose novel interventions [11, 82, 88, 105, 106, 134, 139, 171–173, 177, 198, 202], in particular, motivate their DSCTs as a mean to investigate *new intervention strategies*. Some of these papers investigate similar approaches. The "interaction restraint" proposed by Park et al. [172] and Kim et al. [106], for example, is an intervention mechanism that aims at degrading the interactivity of a device or of a specific service through an unnecessary task to be performed at the beginning of each usage session. Other strategies are instead solutions that have been originally applied for other problems, and that have been transposed to the digital wellbeing context. This is the case of the research of Pinder et al. [177], which seeks to understand whether approach biases for smartphone-addicted users can be reduced through cognitive bias modification, a nonconscious behavior change technology that has been already applied to other domains like healthy eating [176]. Similarly, Park et al. [173] motivate their work by questioning the usefulness of a virtual reality therapy in the field of digital wellness. Interestingly, Lyngs et al. [139] take a step back from what they call "external interventions," e.g., timers and lockout mechanisms, to "shift the focus to how the internal mechanisms of an app can support user agency" (p. 1): redesigning mechanisms like autoplay features and infinite scrolling on mobile applications, indeed, might be more effective than any externally-imposed interventions, according to the authors.

The need of developing novel interventions is also typically associated to a specific user's *context and/or activity* ($N = 6$, [102, 103, 113, 114, 161, 214]). Through the *Let's FOCUS* tool [103], for example, Kim et al. explore "how software-based interventions can be designed and deployed in colleges" (p. 63:2). *Lock n' Lol* [113] and *NUGU* [114], instead, are specifically designed to help group of students studying together.

Six other papers, instead, focus on developing novel intervention solutions that are *generalizable* [80, 127, 160, 166, 167, 186]. The work of Hill et al. [80], for example, analyzes contemporary mobile operating systems to uncover software features that researchers might require to design interventions for moderate Internet use on smartphones of different vendors. Other similar works aim at providing holistic approaches for coping with distractions [186] and overcoming limitations of existing digital wellbeing applications, e.g., excessive "pick & mix" of interventions [166], poor standardization [167], and single-device conceptualizations [160].

3.2.2 Understanding Interventions. Rather than designing novel interventions, a common goal for studying and proposing DSCTs is to *understand* how *interventions* work ($N = 14$, [33, 43, 47, 62, 81, 112, 117, 118, 144, 145, 149, 150, 159, 229]), e.g., to assess their effectiveness on changing users' behaviors. Not surprisingly, such a goal derive from papers included in the *Implemented Tools* corpus, only. Seven papers [47, 144, 145, 149, 150, 159, 229], in particular, analyze the performances of commercially available DSCTs. Mark et al. [149, 150] use *Freedom* [63], a commercial DSCT with which users can set up blocks on their browser, to understand whether blocking distractions is an effective method to improve users' digital wellbeing in real-world settings. Collins et al. [47], instead, investigate whether *RecueTime* [185] has an impact on the users' awareness about their social network usage. Interestingly, the papers of Monge Roffarello and De Russis [159] and Lyngs et al. [144] review a large set of commercially available DSCTs for smartphones and PCs, by searching the Google and Apple app stores, as well as browser extension web stores. Through their first version of the *Socialize* tool, in particular, Monge Roffarello and De Russis assess "in-the-wild" the most popular existing intervention strategies implemented by mobile DSCTs, e.g., timers and lockout mechanisms, with the aim of providing "an overall perspective of contemporary mobile apps for digital wellbeing, and identifying possible issues and opportunities to improve such solutions" (p. 1). Similarly, another work of Lyngs et al. [145] investigates the effectiveness of two popular interventions adopted for reducing Facebook usage on the PC, i.e., removing the newsfeed and goal reminders.

Other papers ($N = 8$, [33, 43, 62, 81, 112, 116–118]) of the *Implemented Tools* corpus present novel DSCTs, but the underlying research goal is the need of understanding how the implemented strategies influence users' behavior. *HabitLab* by Kovacks et al. [116–118], for instance, has been described and evaluated in different studies, e.g., to understand whether rotating browser interventions like timers and persuasive messages increases effectiveness with respect to using static interventions [117] or to understand whether the time saved on some unwanted apps is (or not) redirected to other unproductive activities [116]. *FamilyLink*, instead, has been proposed by Ko et al. [112] to investigate "how participatory parental mediation of smartphone usage by adolescents can overcome restrictive and unilateral mediation approaches" (p. 867). In a similar context, Hiniker et al. [81] developed their *Coco's Videos* platform, thanks to which the same kids can plan their media consumption, to "investigate one possible alternative to today's parental controls and move away from the traditionally authoritarian designs" (p. 10).

3.2.3 Understanding People and Their Usage Patterns. Another common research goal that characterizes the papers included in our *General Corpus* is to understand how *people* use their devices and services, including the same DSCTs (N = 19, [16, 19, 65, 83, 101, 104, 108, 116, 132, 133, 140, 162, 165, 188, 189, 195, 204, 213, 223]). On the one hand, understanding how users benefit from using DSCTs and in what ways is fundamental to inspire future research in this field and develop better solutions for digital wellbeing, e.g., to enhance people's productivity by means of effective feedback [108]. On the other hand, delving into how users use technology is a necessary step to lay the foundation for effective DSCTs.

To investigate how users *use DSCTs* (N = 8, [16, 104, 108, 116, 133, 188, 189, 223], *Implemented Tools* corpus), researchers adopt novel and/or existing implementations of tools for digital self-control. Kim et al. [104], for example, evaluate their *PomodoroLock* tool to explore how office workers would use the provided coercive strategies to cope with distractions. Similarly, Liu et al. [133] present an operating system widget to test "whether providing visual feedback regarding the duration of task suspension can improve primary task resumption" (p. 767). As reported by Rooksby et al. [189], investigating how users interact with DSCTs is also useful to understand which usage data might be of personal interest and value to the user. Providing users with better statistics, in particular, would allow them to gain real-time awareness about how they spend their time online, thus allowing an efficient logging and quantification of the time that is spent using devices like smartphones [16]. Stemming from such an assumption, the aim of the study of Whittaker et al. [223] is to analyze whether real-time awareness provided by the *meTime* tool improves user focus. In addition, some of the papers investigating DSCTs use study the influence of these tools on usage patterns. The work of Kovacs et al. [116], for instance, extends the *HabitLab* tool through a mobile application, and it explores whether the time saved thanks to the DSCT is actually saved or if it is just redirected to other unproductive activities, e.g., usage sessions with mobile apps or websites that are not subjected to any intervention. Monge Roffarello and De Russis [188], instead, propose a DSCT targeting smartphone habits, i.e., recurrent usage patterns that are associated with stable contextual cues. The tool is able to detect such behaviors, and it can assist users to change the habits that are perceived as meaningless.

Research works that analyze *technology use* to extract and propose digital self-control interventions (N = 11, [19, 65, 83, 101, 132, 140, 162, 165, 195, 204, 213], *Qualitative Discussions* corpus), instead, are based on quantitative and qualitative studies, from surveys, e.g., [204], to mixed-method studies composed of semi-structured interviews and log analysis, e.g., [140]. These kind of analysis can help researchers and practitioners to turn their focus from "an overly-determined and prescribed view of technology engagement" to digital self-control tools that promote a more "conscious, mindful, intrinsically-motivated and perhaps even more ethical form of interaction" (Genova et al. [65], p. 1). With this underlying goal, Tran et al. [213] run their study to examine how smartphone users make sense of their habitual phone use, with the aim of finding effective ways to mitigate compulsive smartphone sessions. Similarly, Lukoff et al. [140] conducted their study to understand the smartphone usage patterns that leave people frustrated, while Aranda and Baig [19] explore the dynamics of excessive smartphone use. Besides smartphones, another set of studies explore behaviors change goals (and possible digital interventions) for social media usage, e.g., by exploring how and why people take breaks from social media [195]. Sleeper et al. [204], for example, use behavior-change goals to “explore how people view SNSs as impacting their lives, and to describe the range of goals participants have” (p. 1058). According to the authors, such an analysis is fundamental to explore DSCTs for social networks like Facebook, Instagram, and Twitter. Finally, some papers investigate how people use technology in specific contextual situations, e.g., in the presence of others at home [165] or in caregiving contexts [83], to propose solutions for digital self-control tailored for these specific use cases.

4 INTERVENTIONS FOR DIGITAL SELF-CONTROL AND APPLIED THEORIES

This section reports on the intervention strategies for digital self-control that are proposed and implemented in the papers under analysis, and it summarizes the behavioral theories and constructs that researchers have considered in their studies (**RQ2**). For what concerns strategies, we separately analyzed the implemented interventions, i.e., those included in the *Implemented Tools* corpus, and the proposed interventions, i.e., those discussed in the *Qualitative Discussions* corpus. For the theories and constructs, instead, we analyzed all the papers included in the *General Corpus*.

4.1 Implemented Strategies

Table 5 provides a summary of the tools included in the *Implemented Tools* corpus. For each paper in the corpus, in particular, the table reports a summary of the implemented DSCT, which devices are targeted, and what are the delivery platforms, i.e., the tools with which the DSCTs deliver their interventions to the users. Overall, the 45 analyzed papers include 41 distinct tools for digital self-control, with 3 of them, i.e., *Freedom* [149, 150], *Let's FOCUS* [102, 103], and *HabitLab* (version 1) [117, 118], that are described in different papers. As reported in the table, the majority of these 41 DSCTs (35, 87.50%) use a delivery platform that matches with the target device(s). In particular, 21 DSCTs targeting smartphones [16, 62, 81, 82, 103, 105, 106, 112–114, 127, 134, 159, 166, 167, 171, 172, 188, 202, 223, 229] deliver their interventions through a mobile application, 10 tools targeting PCs are implemented as a browser extension [33, 43, 88, 117, 145, 149, 198, 214] or an application at operating system level [108, 133], while 4 “multi-device” DSCTs [47, 104, 116, 189] include both a mobile app and a computer-based tool, e.g., a browser extension. This means that digital devices are nearly always considered as a source of problems for the user's digital wellbeing but, at the same time, they are also the mean with which the interventions against these problems are delivered to the user. There are, however, some notable exceptions. The software of *Knob* [186], for instance, comes with a tangible object to visualize real-time feedback about user's current availability. Similarly, *Time Machine* [11] has been developed as an ambient display and tangible interface for time management distinctly separated from the task environment, while *Crank That Way* [205] is a tangible system that requires users to continuously turn a hand crank in order to use Twitter. Pinder et al. [177], instead, use a trial to be performed on a tabletop to teach users how to reject unconscious and repeated smartphone usage sessions. Morley et al. [161] and Park et al. [173] developed their interventions against dangerous smartphone behaviors and online gaming addiction, respectively, in the form of a virtual reality experience. Finally, Kolling et al. [115] recently developed a community-driven framework, named *GreaseDroid*, to allow non-expert users to modify their own mobile apps through customized patches, with the aim of alleviating dark patterns.

Table 5. Datasheet excerpt of all 45 papers included in our *Implemented Tools* corpus including summary, target, and delivery platform of the implemented DSCTs. Overall, the 45 papers include 41 distinct DSCTs.

Reference	Summary	Target	Delivery Platform
Riedl et al. [186]	*Knob*, a tangible controller for multiple distraction blocking mechanisms to tackle social and digital distractions.	PC, smartphone	Tangible object
Shen et al. [202]	*APP*, a mobile application that provide users with alerting and reminding information based on device usage statistics.	Smartphone	Mobile app
Kim et al. [105]	*GoalKeeper*, a mobile application that locks the user into the self-defined daily use time limit with restrictive intervention mechanisms.	Smartphone	Mobile app
Lyngs et al. [145]	Two browser extensions implementing two interventions to Facebook overuse, i.e., goal reminders and removing the newsfeed.	Social networks	Browser extension
Pinder et al. [177]	A novel experiment applying cognitive bias modification (CBM-Ap) techniques on a Tabletop to counter smartphone overuse.	Smartphone	Tabletop
Kim et al. [106]	*LocknType*, a mobile application that make participants complete lockout tasks before opening distractive applications.	Smartphone	Mobile app
Tseng et al. [214]	*UpTime*, a browser extension and chatbot support workers' transitions from breaks back to work through time limits and proactive dialogues.	PC	Browser extension, chatbot
Monge Roffarello et al. [159]	*Socialize* (version 1), a digital wellbeing mobile application integrating self-monitoring statistics, timers, and lockout mechanisms to counter smartphone overuse.	Smartphone	Mobile app
Kovacs et al. [116]	*HabitLab* (version 2), a Chrome extension and a mobile application through which the user can choose to set up different interventions to specific web sites and/or apps.	PC, smartphone	Browser extension, mobile app
Kovacs et al. [117]	*HabitLab* (version 1), a Chrome extension that let users set up limit goals for specific websites and then rotate a number of interventions to assist users in fulfilling their goals while reducing attrition.	PC	Browser extension
Okeke et al. [166]	*Good Vibrations*, a mobile application that combine nudge theory and negative reinforcement through a repeating phone vibration that spur the user to stop using a target app.	Smartphone	Mobile app
Okeke et al. [167]	A framework for conducting behavior change research studies using smartphones in-the-wild.	Smartphone	Mobile app
Mark et al. [149]	*Freedom*, a commercial DSCT that can block access to specific websites or to the entire Internet.	PC	Browser extension
Kim et al. [104]	*PomodoLock*, a mobile application and a browser extension implementing Pomodoro sessions during which the user can block the access to specific apps and/or websites.	PC, smartphone	Browser extension, mobile app
Kim et al. [103]	*Let's FOCUS*, a mobile application with which users can enter a "virtual room" where specific apps are blocked and notifications are muted.	Smartphone	Mobile app
Mark et al.[7] [150]	*Freedom*, described above.	PC	Browser extension
Kim et al.[8] [102]	*Let's FOCUS*, described above.	Smartphone	Mobile app
Chisler et al. [43]	A browser extension that allow users to set up time limits for online streaming services like Netflix.	PC	Browser extension
Foulonneau et al. [62]	*TILT*, a mobile application that monitors smartphone usage and delivers persuasive messages to the user.	Smartphone	Mobile app
Andone et al. [16]	*Menthal*, a mobile application that summarizes the overall smartphone usage into a single number, i.e., the "MScore."	Smartphone	Mobile app
Morley et al. [161]	An immersive user experience based on virtual reality with which participants experience the negative consequences of smartphone use while driving.	Smartphone	Virtual Reality
Ko et al. [113]	*Lock n' LoL*, a mobile application that allows a group of people to mute notifications and lock the usage of the smartphone to perform a group activity.	Smartphone	Mobile app
Kim et al. [108]	*TimeAware*, a widget at operating system level which displays time spent on "distracting" or "productive" programs.	PC	OS widget
Hiniker et al. [82]	*MyTime*, a mobile application that monitors the time spent and number of visits on specific apps by allowing users to set up daily time limits.	Smartphone	Mobile app
Whittaker et al. [223]	*meTime*, a PC application that shows users how they allocated their time across applications within the last 30 minutes.	Smartphone	Mobile app
Rooksby et al. [189]	*ScreenLife*, a multi-device personal tracking system that enables users to collect and view data about the use of their digital devices.	PC, smartphone	Mobile app, OS app, website
Ko et al. [112]	*FamiLync*, a mobile application with which members of a family can monitor smartphone usage together and set up daily usage limits.	Smartphone	Mobile app

[7] This is a preliminary paper version of [149].
[8] This is a preliminary paper version of [103].

Table 5 Continued

Reference	Summary	Target	Delivery Platform
Collins et al. [47]	*RescueTime*, a commercial DSCT which provides visualisations of how much time is spent in different mobile applications and websites.	PC, smartphone	Browser extension, mobile app
Ko et al. [114]	*NUGU*, a mobile application that allows users to set up their own goals for limiting smartphone usage and to share their achievements with friends.	Smartphone	Mobile app
Ahmed et al. [11]	*Time Machine*, an ambient display and tangible interface for time management that uses colored glass marbles to represent units of time and clear cylinders as tasks.	PC	Tangible object
Löchtefeld et al. [134]	*AppDetox*, a mobile application with which users can create rules to avoid using specific apps.	Smartphone	Mobile app
Liu et al. [133]	Two different visual feedback that indicate how much time the user has been away from her main task.	PC	OS widget
Park et al. [173]	A virtual reality therapy for online-gaming addiction consisting of a four-week treatment period.	Online gaming	Virtual Reality
Ahn et al. [127]	*SAMS*, a mobile application that allows users to track their smartphone usage and set up time limits for using specific apps.	Smartphone	Mobile app
Park et al. [172]	A mobile application that forces users to perform unnecessary cognitive tasks before opening specific apps.	Smartphone	Mobile app
Monge Roffarello et al. [188]	*Socialize* (version 2), a mobile application that detects habitual usage patterns and allows users to change these behaviors via implementation intentions.	Smartphone	Mobile app
Borghouts et al. [33]	*TimeToFocus*, a browser extension that allows users to select a main task window to focus on and visualizes a notification showing how long on average users switch away from the task window.	PC	Browser extension
Hiniker et al. [81]	*Coco's Videos*, a video-viewing platform for preschoolers designed to support them in learning to self-manage their media consumption through limits and UI's modifications.	Smartphone	Mobile app
Park et al. [171]	*GoldenTime*, a mobile app that promotes self-regulated usage behavior via system-driven proactive timeboxing and micro-fnancial incentives framed as gain or loss for behavioral reinforcement.	Smartphone	Mobile app
Schwartz et al. [198]	*Time Sidekick*, a prototype that implements four novel design patterns for reducing risk of abandonment of DSCTs.	PC	Browser extension
Song et al. [205]	*Crank That Feed*, a system for using Twitter that requires users to continuously turn a hand crank to power their social media screen.	Social networks	Tangible object
Zhou et al. [229]	A built-in gaming gradual intervention system (G-GIS) designed to help adult gamers achieve their desired gaming habits in an autonomous and acceptable manner.	Online gaming	Mobile app
Inie et al. [88]	*Aiki*, a browser extension designed to redirect a user to a learning platform for a fixed amount of time before accessing distractive websites.	PC	Browser extension
Kovacs et al. [118]	*HabitLab* (version 1), described above.	PC	Browser extension
Kollnig et al. [115]	*GreaseDroid*, a community-driven app modification framework enabling non-expert users to disable dark patterns in mobile apps selectively.	Smartphone	Community-driven framework

To further dig into the functionality of the analyzed DSCTs, we applied the coding scheme adopted by Lyings et al. [143] (Figure 3), by classifying the implemented interventions in *block/removal*, *self-tracking*, *goal-advancement*, and *reward/punishment* features.

Block/Removal. By analyzing the functionality of the DSCTs included in our corpus, we coded a *block/removal* feature 57 times, with each analyzed tool that includes on average 1.43 features of this kind (SD = 1.43). The most common features in this group are the possibility of setting up a *time limit* on how much time can be spent on a given technological source (N = 13, [81, 82, 103, 105, 112, 116–118, 127, 134, 159, 171, 229]), and the possibility of inserting an access *block* to a device, an app, and/or a website that is considered as a source of digital wellbeing problems (N = 18, [43, 47, 82, 103–105, 112–114, 116–118, 127, 134, 149, 159, 186, 214]). Less used features in this category include, between others, *feature minimisation* (N = 6, [81, 115–118, 145]), e.g., removing the Facebook newsfeed [145] or removing autoplay features in video-viewing platforms [81] and the proactive insertion of irrelevant *effortfull tasks* (N = 5, [106, 145, 172, 205, 214]), e.g., the "interaction restraint" approach adopted by Park et al. [172] and Kim et al. [106]. Another small but interesting set of DSCTs can insert a

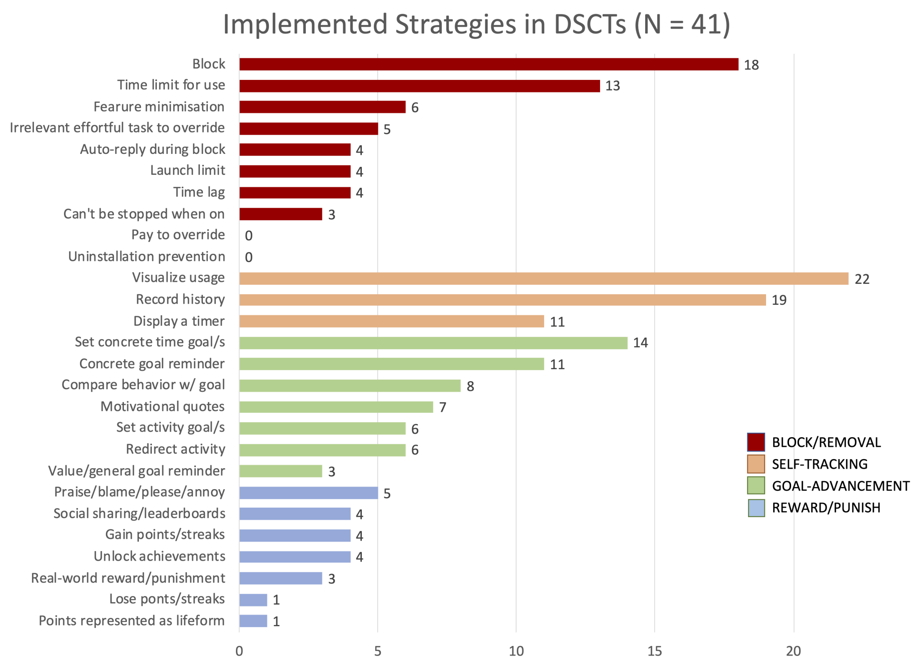

Fig. 3. The functionality implemented by the DSCTs under analysis (*Implemented Tools* corpus) according to the categorization of Lyngs et al. [143].

time lag for the usage of the involved device and/or service (N = 4, [116–118, 198]). Such a strategy does not block the usage of websites or mobile apps, but just makes them load more slowly to discourage continuous usage. The *Time Sidekick* browser extension [198], for example, implements two continuously variable interventions that adds a delay before the page loads and further delays to dynamically loaded content in the page, respectively. None of the analyzed DSCTs, instead, adopt a mechanism against the *uninstallation* of the tool. This can be explained by the characteristics of the operating systems hosting the DSCTs, e.g., Android, that typically do not allow these mechanisms unless the user has root privileges [159].

Self-Tracking. Another prevalent set of features, coded 52 times in total, is the one related to *self-tracking*. This confirms the self-monitoring nature of contemporary DSCTs [159], that are typically designed to track user's behavior and provide feedback. To this end, 19 DSCTs in our corpus [16, 33, 47, 62, 88, 103–105, 108, 112–114, 116–118, 159, 171, 198] persistently *record the history* of the user with a device and/or a service, e.g., by saving statistical data about usage sessions, and 22 tools [16, 47, 62, 88, 103–105, 108, 112–114, 116–118, 133, 159, 171, 188, 189, 198, 202, 223] *visualize usage data* to the user. Visualization may include historical data as well as real time feedback, e.g., those provided by the *Knob* tangible device [186] about the worker's availability. Usage data is also typically summarized in real time through the usage of a *timer* (N = 11, [33, 43, 62, 82, 103, 105, 116–118, 171, 214]).

Goal-Advancement. Several analyzed DSCTs also include one or more *goal-advancement* feature (55 overall mentions), although the most common feature of this category, i.e., the possibility of setting up a *concrete time goal* for using a technological source for a given amount of time, is present in 14 DSCTs, only [11, 43, 47, 82, 105, 112, 114, 116–118, 166, 167, 202, 229]. Besides allowing users to set up these goals, the majority of these tools (N = 11, [43, 47, 82, 105, 113, 114, 116–118, 145, 167]) also employ *goal reminders*, e.g., by displaying pop-ups when the user has exceeded a defined time threshold, while 8 tools [43, 47, 103, 113, 114, 116–118] allow users to *compare* their actual *behavior* with their defined goals, e.g., by visualizing usage statistics. Less used interventions in this category include the possibility of setting up more generic goals, e.g., related to specific *users' activities* (N = 6, [82, 112, 114, 145, 188, 202]), the usage of *motivational quotes* (N = 8, [47, 62, 113, 116–118, 186]), and automatic *redirections* (N = 6, [82, 88, 116–118, 188]), e.g., to spur the user to use a more productive service or activity (see the Aiki tool [88], which redirects users to a platform for foreign language learning before letting them access a distractive website).

Reward/Punishment. Overall, we coded *reward/punishment* features 22 times. Only a small portion of the analyzed DSCTs (N = 4, [103, 112–114]), typically those developed for students, include the possibility of *sharing* the own achievements with others (see *NUGU* [114] and *Lock n' Loll* [113]). When users reach an achievement (or fail to meet a goal), some DSCTs (N = 5, [47, 116–118, 166]) can *praise* (or *blame*) them for their behavior. Some other tools employ a gamification approach by allowing their users to *gain* (N = 4, [112, 114, 133, 202]) or *loose* (N = 1, [133]) points. Such points, sometimes represented through a *lifeform* metaphor [202], can also be used to *unlock achievements* (N = 4, [116–118, 229]). Interestingly, a small set of DSCTs (N = 3, [103, 171, 229]) also adopt *real world rewards/punishments*. The Gaming Gradual Intervention System (G-GIS) analyzed by Zhou et al. [229], for example, offers a free audio book to its users when they take a break from the gaming platform. The *GoldenTime* mobile app, instead, promotes self-regulated usage behaviors through micro-financial incentives framed as gain or loss for behavioral reinforcement.

As the reported classification suggests, the DSCTs included in the *Implemented Tools* corpus are (nearly) always "self-programmed" by users, without any particular support by the tool. By interpreting their usage statistics and/or their usage self-perception, in particular, users had often to decide which websites or mobile apps need an intervention, as well as how the intervention should operate, e.g., which is the time limit for a given app. Noteworthily, the second version of *Socialize* [188] is the only example of a DSCT that can learn from smartphone user's data and provide proactive support to define interventions. The tool constantly monitors the user's behavior with their mobile device, and it adopts a machine learning methodology, based on association rule mining, to detect "smartphone habits," i.e., recurrent usage patterns associated with stable contextual cues like locations and time. Stemming from such a monitoring process, *Socialize* proactively notifies the detected smartphone habits in real-time, and it allows the activation of personalizable just-in-time reminders that encourage the user to avoid the identified behaviors when they happen again.

To understand whether our classification is consistent with previous reviews of DSCTs, we also compared the features of the DSCTs in our corpus (Figure 3) with the features that have been found by Lyings et al. in their review of 367 DSCTs available on popular app and web stores (see [143], Figure 3, p. 7). We found that, while the different sizes of the analyzed samples result in some minor differences within each group of features, the overall distribution is quite similar. As in [143], for instance, we found that the majority of our DSCTs include interventions to block or remove distractions, while only a small portion of them reward or punish their users to solidify a behavior or discouraging it, respectively. This highlights a promising interconnection between the digital wellbeing seen as a research field and its real-world application. Indeed, some DSCTs that have been developed as a result of a research study, e.g., *HabitLab* [116, 117], are also available to the general public on app and web stores, and they were therefore also included in the study of Lyngs et al. [143]. On the contrary, researchers are also using commercial DSCTs to study whether and how these tools influence people's behavior (see the work of Collins et al. [47] with *RecueTime*, and the two papers of Mark et al. [149, 150] with *Freedom*).

4.2 Proposed Strategies

We separately analyzed the papers included in the *Qualitative Discussions* corpus to investigate what strategies researchers are suggesting, e.g., to inform future works and possibly improve contemporary DSCT implementations. Figure 4 shows a classification of these proposed strategies extracted from the 17 papers under analysis.

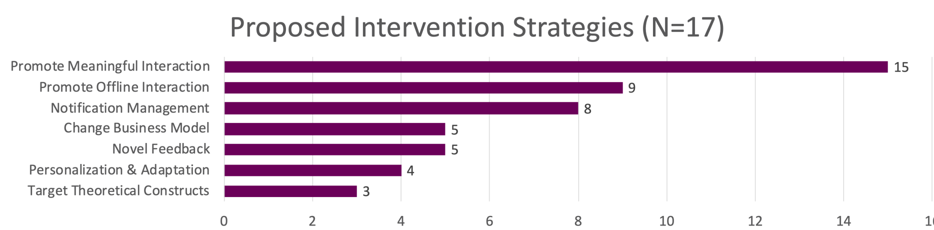

Fig. 4. The strategies proposed by the research works included in the *Qualitative Discussions* corpus. The majority of the 17 analyzed papers include more than one proposed strategy.

As the chart shows, the majority of the proposed strategies focus on either *promoting meaningful interactions* (N = 15, [65, 83, 139, 140, 160, 162, 195, 204, 213, 224]) or *promoting offline interactions* (N=9, [19, 64, 65, 83, 144, 160, 204]). These proposals deviate significantly from the majority of the implemented interventions described in Section 4.1, in which the goal is primarily to block (or mitigate) the usage of a given device, app, or website.

Promoting meaningful interaction, in particular, means moving beyond lock-out mechanisms [160, 213] by guiding users toward (digital) experiences they will find meaningful [213], e.g., increasing interactions on social networks that users consider beneficial [204]. As suggested by Genova et al. [65], there is the need to strive for alternatives instead of avoidance, by promoting joyful opportunities for engagement. In the smartphone context, for example, the work of Lukoff et al. [140] suggests researchers to explore the design of meaningful experiences, by taking into account how mobile devices respect users' autonomy. According to the authors, a DSCT could detect when the smartphone is/is not being used with intention, and it could dynamically change its interface accordingly, e.g., to encourage users to move on when their original purpose is achieved. Similarly, Lukoff et al. [139] call researchers to design DSCTs that support microplanning, e.g., "by asking the user to review and refect upon their past usage data and develop a plan for their use over the next month" (p. 13). Tran et al. [213], instead, suggest the exploration of designs and metrics that are tailored to the current feelings and contextual situations of the user, e.g., moments of downtime or socially awkward situations. Designing for meaningful use is obviously not straightforward: it requires researchers to balance users' desire for free will with the constraints of the involved app or website [195] and, at the same time, to redefine metrics for users' satisfaction, e.g., to detect how meaningful a digital experience is likely to be and how much time the user is going to spend [213]. Furthermore, interventions should have different levels of control depending on the underlying usage intention of the user, to encouraging intentional sessions while discouraging passive use [139].

Instead of promoting meaningful "digital" usages, *promoting offline interactions* is focused on facilitating disconnection [19] through the suggestions of possible "offline" alternatives [65]. This could be achieved in different ways and through different strategies. Genç and Coskun [64], for example, envision a DSCT that is able to boost the current conversation of a group of users by collecting data about their interests and proposing new topics to discuss. The work of Aranda and Baig [19], instead, proposes several alternative solutions to design interventions that promote offline interactions, from using media outreach campaigns that increase general awareness of screen overuse, to DSCTs that allow users to find support partners. Independently of the suggested strategies, the ultimate goal of these proposals is to help users to practice and form new "healthier" habits that can replace compulsive and problematic digital behaviors [19, 144].

Other research works in our *Qualitative Discussions* corpus are more aligned with contemporary DSCT implementations, and try to propose improvements for a better *notification management* (N = 8, [64, 101, 132, 165]). These proposals include smarter filters able to analyze notifications content [64, 101, 132], context-aware notification systems [101, 165], and processes to make the behavior of checking notifications more difficult, e.g., via gamification [64]. Similarly, researchers also discuss how to make existing interventions for digital self-control more *personalized and adaptable* (N = 4, [160, 165, 195]), e.g., to blend alerts and notifications to the users' temporal rhythms and home/work boundaries [165], and propose strategies to *target specific theoretical constructs* (N = 3, [132, 144]) and design *novel feedback* (N = 5, [64, 65, 162, 165]). Some of these novel feedback could be reasonably implemented in contemporary DSCTs without any particular effort. Genç and Coskun [64], for example, suggest that DSCTs could emphasize the value of time by informing users about what they can do in real life with the time they spend on their smartphones, while Moser et al. [162] state that mobile phones should use their notification leds to promote awareness of social activities, e.g., to indicate the nature of the use to people nearby. Other novel feedback are instead more futuristic. This is the case, for example, of a smartphone that "changes its color or heat according to the frequency of its use [64]", or that "grows old and wrapped by an ivy as the quality of the interaction increases."

Finally, researchers (N = 5, [80, 139, 224] highlight possible ways to *change the business model* of tech companies, which nowadays incentivizes continuous and frequent usage in order to optimise advertising revenue [144]. Lukoff et al. [139], for example, highlight the need of rethinking what "relevance" means for recommeder systems. They state, in particular, that developers of these technologies should "first take into account the problem of when to show recommendations, before moving on to the local problem of which items to recommend" (p. 12). At the same time, the authors propose to develop a common language of harmful designs that lead to attentional harms, with the aim of motivating key stakeholders to support designs that promote users' sense of agency. With similar goals, Hill et al. [80] ask for mobile operating systems that are more open to the community, e.g., by improving API transparency and by creating permissions to collect usage data.

4.3 Adopted Behavioral Theories and Constructs

We examined all the 62 papers included in our *General Corpus* to understand whether they *explicitly based* their implemented or proposed interventions on any behavioral theory and/or specific construct. Within behavioral science, a *behavioral theory* seeks to systematically understand events or situations, and can be defined as "a set of concepts, definitions, and propositions that explain or predict these events or situations by illustrating the relationships between variables" (Glanz and Rimer [68], p. 4). *Constructs*, instead, are the fundamental components of a behavioral theory, and can be defined as "the basic determinants or mechanisms that a theory postulates to influence behavior" (Hekler et al. [77], p. 3309). Overall, our theoretical analysis, reported in Table 6, aims to extend the preliminary investigation done by Lyngs et al. [143] over a smaller set of papers (N = 17) that "were not collected through a formal systematic review process" (p. 13:3).

Table 6. The behavioral theories and constructs that the 37 papers under analysis mention when describing the design and the evaluation of the proposed DSCTs.

Theory	Summary	Paper/Authors
Dual system theory [93, 207]	Dual system theory defines two mental processes that compete to determine behavior: System 1 processes (fast, nonconscious, associative) and System 2 processes (slower, deliberative, conscious).	Lyngs et al. [145]; Kim et al. [106]; Schwartz et al. [198]; Lyngs et al. [144];
Goal setting theory [135, 136]	Goal-setting theory refers to the effects of setting specific and sufficiently difficult goals that are accepted by individuals on subsequent performance.	Kovacs et al. [117]; Hiniker et al. [82]; Collins et al. [47]; Sleeper et al. [204];
Operant conditioning [206]	Operant conditioning refers to the process of learning shaped by rewards and punishments for a behavior. Through operant conditioning, an association is made between a behavior and a consequence for that behavior, be it negative or positive.	Kovacs et al. [117]; Okeke et al. [166]; Liu et al. [133]; Park et al. [171];
Social cognitive theory [21]	Social cognitive theory states that the interaction between the environment, the cognitive and affective characteristics of the individual, and her existing behaviors, determine how the individual will behave. According to the theory, in particular, knowledge acquisition can be related to the observation of others within the context of social interactions, experiences, and external influences.	Ko et al. [112]; Ko et al. [114]; Park et al. [171];
Transtheoretical model [182]	The transtheoretical model identifies six stages characterizing behaviour change, from precontemplation to active modification, and ten processes targeting self-efficacy and decisional balance to move between these stages.	Kim et al. [105]; Park et al. [171];
Fogg behaviour model [61]	The Fogg behaviour model is a general cross-domain model according to which motivation, ability, and a trigger must all be present at the same time so that a particular behavior can be performed. Motivation and ability, in particular, define a space where both a behavior and the resistance to change it can be characterized.	Foulonneau et al. [62]; Sleeper et al. [204];
Self-determination theory [56]	The self-determination theory is a macro theory of human motivation and personality that studies the motivation behind choices people make without external influence and interference. It focuses on the degree to which human behavior is self-motivated and self-determined.	Hiniker et al. [81]; Lukoff et al. [139];
Uses and gratification theory [97]	Uses and gratification theory is a tool to investigate why and how people actively use specific media to meet their needs. It was originally developed to study the influence of mass media, e.g., television, but it has been recently applied to understand the usage of digital media like social networks and smartphones.	Kim et al. [106]; Lukoff et al. [140];
Expectancy-value theory [90]	The expectancy-value theory suggests that expectations as well as values or beliefs affect subsequent behavior. According to the theory, people evaluate their interest and attainment by considering utility and cost.	Kim et al. [106];
Nudge theory [209]	Nudge theory describes interventions that direct people to perform a particular behavior without eliminating their freedom of making the final choice. It is based on positive reinforcement as a way to influence behavior and decision making.	Okeke et al. [166];
Theory of planned behavior [12]	The theory of planned behavior states that intentions drive people's behaviors. According to the theory, attitude, subjective norm, and perceived behavioral control, together shape an individual's behavioral intention, which in turn determine the subsequent behavior.	Kim et al. [106];
Theory of reasoned action [14]	The theory of reasoned action studies the links between an intention to act in a certain way, the individual's attitudes, and subjective norms, by predicting that a behavioral intent arises from the individual's evaluation and strength of a belief, and her motivation to comply the expectations of others.	Kovacs et al. [117];
Theory of distributed cognition [86]	The distributed cognition theory seeks to understand the organization of cognitive systems by extending the cognitive concept beyond the individual. According to the theory, indeed, knowledge is also related to the individual's social and physical environment.	Ahmed et al. [11];
Elaborated intrusion theory of desire [98]	The elaborated intrusion theory of desire focuses on the roles of intrusive thoughts and elaboration of multisensory imagery. It characterizes the conscious experience of desiring something as a cycle of mental elaboration that stems from an initial intrusive thought.	Park et al. [172];
Behavioral economics [95]	Behavioral economics studies how different factors, from psychological statuses to cultural and social factors, affect the decisions made by individuals and institutions. It provides a framework to dig into when and how people make systematic errors or biases.	Park et al. [171];
Media effects theory [35]	Media effects theory explain how the mass media influence the attitudes and perceptions of the users.	Schoenebeck [195];
Stimulus-organism-response model [152]	The stimulus-organism-response model states that environmental stimuli (S) first evokes an emotional reaction (O) in persons, which in turn leads to either approach or avoidance behavioral responses (R). It aims at explaining the consumer decision making process.	Lin et al. [132]
Flow theory [51, 52]	The flow theory describes a "state in which people are so involved in an activity that nothing else seems to matter; the experience is so enjoyable that people will continue to do it even at great cost, for the sheer sake of doing it [52]."	Lin et al. [132]

Table 6 Continued

Theory	Summary	Paper/Authors
Behavior change wheel taxonomy [157]	The behaviour change wheel is a framework to identify relevant interventions starting from the description of the target behaviour. It uses the COM-B model, and it was developed through a systematic review of 19 existing behavior change frameworks.	Kovacs et al. [116];
Social proof [203]	Social proof is is a psychological and social phenomenon that suggests that people who does not know what the proper behavior for a certain situation is will look to other people to copy what they are doing in an attempt to undertake the right behavior.	Kovacs et al. [117];
Self-consistency [126]	Self-Consistency focuses on the importance of the self in the regulation of thoughts and ideas. Self-consistency is a motivational component of the self-concept that characterize many different theories. It postulates that people are motivated to act in a way that is congruent with their understanding of themselves.	Kovacs et al. [117];
Status quo bias [190]	The status quo bias is an emotional bias that characterize a preference towards the current state of affairs, when any change from the current baseline, i.e., the status quo, is perceived as a loss. Studies demonstrated that decision makers often exhibit a significant status quo bias.	Kovacs et al. [117];

We found that only 24 papers out of 62 explicitly refer to behavioral theories. As reported in Table 6, such papers mention 18 behavioral theories (from *dual system theory* [93, 207] to the *flow theory* [51, 52]), 1 framework that includes several theories (the *behavior change wheel taxonomy* [157]), and 3 constructs that are key determinants of different theories (*social proof* [203], *self-consistency* [126], and *status quo bias* [190]). Theories and constructs are used to inform the proposal and design of novel interventions [11, 62, 82, 106, 112, 114, 117, 133, 166, 171, 172, 198], to characterize and select existing interventions to be evaluated [47, 81, 116, 118, 144, 145, 229], to analyze how users use digital devices and services [132, 139, 140, 195, 204], and even to select participants for evaluating a DSCT (see Kim et al. [105] and Park et al. [171], who used the *transtheoretical model* to restrict study participation to users that would like to reduce smartphone use). The remaining 38 papers do not specify any underlying theory nor construct, but rather propose interventions or describe DSCTs that have been informed uniquely by empirical evidence, e.g., through surveys, interviews, or reviews of existing apps. Only 4 theories are explicitly mentioned in at least 3 papers: *dual system theory*, *goal setting theory*, *operant conditioning*, and *social cognitive theory*.

Dual System Theory. Dual system theory [93, 207] argues that behavior is determined by two distinct mental processes, i.e., System 1 and System 2. System 1 represents nonconscious, automatic response heuristics that rapidly emerge as a result of an exposure to a given cue, e.g., a particular contextual situation. System 2, instead, represents conscious, deliberative responses that are goal-directed, e.g., via a specific intention of the user. According to dual system theory, in particular, behavior emerges as the result of an interplay between System 1 and System 2. Most behavioral reactions, for instance, begin with System 1, e.g., when individuals lack cognitive capacity or working memory [212], but can be influenced by conscious System 2 goals [93]. On the contrary, System 2 processes may also become controlled by System 1 over time [94]. Dual system theory is mentioned by 4 papers [106, 144, 145, 198] in our *General Corpus*. Lyngs et al. [145] adopt a dual system model of self-regulation [143] to categorize and select two different existing interventions against Facebook overuse to be evaluated. According to the authors, removing the Facebook newsfeed prevents unconscious behaviors triggered by the newsfeed (System 1), while a goal advancement intervention enables System 2 control by keeping the goal in the user's working memory. Similarly, Schwartz et al. [198] stem from the dual system theory to define their novel design patterns for reducing risk of abandonment in DSCTs. Kim et al. [106], instead, state that the "interaction restraint" intervention included in their *LockType* tool facilitate mindful interaction [49], since it is able to create a gulf of execution on gratification seeking that can encourage users to switch from System 1 to System 2 thinking. Finally, Lyngs et al. [144] apply a dual systems model of self-regulation to organize and evaluate the design features of a large set of existing DSCTs.

Goal Setting Theory. Goal setting theory [135, 136] explores how conscious and explicit goals drive behavioral repetition and support habit formation. To be effective, these goals must be accepted by the user, and they must be sufficiently *specific* and *hard* to be accomplished in order to influence behavior. Previous work [135], indeed, demonstrates that difficult goals are more effective than vague objectives, especially when they are self-defined and not externally imposed [137], e.g., by the DSCT. While 20 DSCTs included in our corpus allow users to set up goals (see Figure 3), only 4 papers [47, 82, 117, 204] in our *General Corpus* explicitly mention the goal setting theory, e.g., as the self-regulation foundation which inspired the design [82, 117] and the evaluation [47] of interventions. In these papers, goals refer to how much time a user would like to spend on specific websites (see *HabitLab* [117]) or mobile applications (see *MyTime* [82]), and are always self-defined by the user.

Operant Conditioning. Operant conditioning [206] is part of the behaviourism approach [40], which postulates that behavior is learned from interacting with the environment through stimulus–response associations that are learned via repetition. Within such an approach, operant conditioning characterizes the connection between a stimulus-response link and a positive reward for a wanted behavior (or a negative outcome for an unwanted behavior). Rewarding a behavior strengthens the link between the stimulus and the response, thus increasing the possibility of performing the same behavior in the future. On the contrary, punishing a behavior aims at reducing the strength of the stimulus-response link. We found an explicit mention of operant conditioning in 4 papers [117, 133, 166, 171], all of them coming from the *Implemented Tools* corpus. In their *Good Vibratios* tool, Okeke et al. [166] implement a negative reinforcement strategy, by "punishing" the user through phone vibration when the usage of a mobile app surpasses a time threshold. Liu et al. [133], instead, evaluate two persuasion strategies based on punishment, i.e., visualizing a withering flower, and reward, i.e., visualizing a blooming flower. Similarly, Park et al. [171] explore the effectiveness of using micro-financial incentives, framed as loss or gain, to reinforce "small changes," e.g., for a given hour, use your phone less than 10 minutes. Finally, the first version of the *HabitLab* tool [117] includes two interventions that delay or remove a reward, i.e., making the user wait 10 seconds before visiting Facebook and hiding the Facebook newsfeed, respectively. While also Lyngs et al. [145] adopt similar interventions for Facebook, they characterize them through the lens of the dual system theory rather than operant conditioning (see the next paragraph).

Social Cognitive Theory. Social cognitive theory [21] states that behavior is learned through the observation of others within the context of social interactions, experiences, and external influences. A desired behavior, in particular, is determined by the interaction between the environment, the cognitive and affective characteristics of the individual, and her existing behaviors, and it is more likely to be performed when environmental barriers are low, and the user's self-efficacy is high [20]. Overall, 3 papers [112, 114, 171], all of them in the *Implemented Tools* corpus, explicitly mention the social cognitive theory as the theoretical basis of the proposed interventions. *NUGU* [114], for example, is designed to promote social learning and competition between students that can share their limiting practices with others. *FamilyLink* [112], instead, allows parents and children to limit device usage together. To this end, the tool offers a virtual space to foster social awareness and improve self-regulation.

Table 7. The challenges reported in the 62 papers of our *General Corpus* that characterize the discussion, the design, and the evaluation of DSCTs.

General Themes	Challenges
Adaptability (N=7)	Adapting the interventions to the target device and/or service, since each device, app, or website has its own type of usage, e.g., a device can be personal or shared with others.
	Adapting the interventions to the current users' needs, preferences, and usage patterns, e.g., to fit different interventions to various users in different contexts or to keep users informed of their device usage while minimizing notification overload.
	Adapting the interventions to the different levels of physical capability and technology competence of the users, e.g., to take into account differences between parents and children in parental-control DSCTs.
Accuracy (N=20)	Accurately detecting the current contextual situation in which the user is using her devices, as well as the underlying usage intention and feelings. Devices, for instance, can be used both as entertainment and work tools, and "work time" and "free time" are not necessarily distinct.
	Avoiding intervention bugs and incorrect statistics. An intervention delivered at inappropriate time or a wrong visualized statistic about device use may lead the user in abandoning and/or mistrust the tool.
	Helping users to understand the visualized statistics, e.g., through sufficient and accurate details, as well as providing information about why an intervention has been delivered. Current visualizations in DSCTs are often difficult to interpret, since such tools are typically used in a retrospective manner.
Technical (N=18)	Overcoming platforms restrictions to design cross-platforms DSCTs. Some platforms like Apple iOS are more restrictive than others, e.g., Android, since they do not allow developers to access sensitive information such as phone usage statistics. Consequently, it is challenging to implement the same intervention strategies across devices with different operating systems.
	Developing DSCTs that target multi-device usage. The majority of contemporary DSCTs target a specif device, e.g., the PC. Without taking into account the different devices owned by a user, the same user can easily circumvent a DSCT by using another device, e.g., the smartphone.
	Mitigating software aging and regular updates. DSCTs are particularly vulnerable to software aging, i.e., natural degradation of software capabilities, and to changes in the platforms that host them, e.g., because they become intentionally inhospitable.
Attrition (N=6)	Motivating the user to to keep using the DSCT in order to change her behavior while minimizing attrition, i.e., the tendency of users to abandon the tool after having using it for some time.
Evaluation (N=12)	Measuring the effectiveness of DSCTs through long-term evaluations and follow-up periods. This is particularly required since behavior change is a long and complex process: current short-term studies can only speculate about the effectiveness of DSCTs over long period of time.
	Going beyond self-reported measures to consider more complex measures. Self-report, indeed, often correlates poorly with actual use of digital devices, while measuring a user's sense of meaningfulness, for example, is undoubtedly harder than tracking time spent in an app.
	Designing studies that allow to capture the real effect of the evaluated DSCTs. Existing DSCTs, in particular, often combine a variety of different interventions at the same time, making it difficult to understand what behavioral theory or technique contributes more to the success of the tool.
Business Model (N=8)	Dealing with the business model of contemporary tech companies, also called the "attention economy," which incentivises the design of technology that promotes frequent and continuous usage. The usage of DSCTs, indeed, may hinder business profitability of key stakeholders.

5 CHALLENGES AND ETHICS IN DEVELOPING FOR DIGITAL SELF-CONTROL

This section delves into the challenges that characterize the proposal, implementation, and/or evaluation of DSCTs, as well as the extent to which researchers describe ethical challenges or implications in their works on digital self-control (**RQ3**). To run this analysis, we analyzed all the papers included in our *General Corpus*.

5.1 The Challenges of Developing For Digital Self-Control

Table 7 summarizes the research challenges highlighted by the researchers in their papers. We coded challenges in terms of *interventions adaptation*, *interventions accuracy*, *technical constraints*, *attrition*, *evaluation*, and *business model*.

5.1.1 Adapting Interventions to Devices, Users, and Usage Patterns. A challenge that characterizes the development of DSCTs regards the *adaptability* of the implemented interventions (N = 7, [83, 105, 112, 132, 189, 214, 224]). As reported in the following, such an adaptability involves different aspects. Researchers [132, 189] highlight that interventions should adapt to the *target device* or *service*. This is particularly important since each device has its own type of usage. Differently from personal devices like smartphones, for instance, tablets can be shared with other members of the family [189]. Even usage sessions are intrinsically different between mobile devices, with smartphones that are used more often than tablets but with shorter usage sessions [84]. Furthermore, while portable devices are typically used for a variety of tasks, from chatting to watching films, other devices have a more specific type of usage, e.g., desktop PCs for working and smartwatches for tracking sport activities [160]. Besides adapting to different devices, other 4 papers [83, 105, 214, 224] underline that interventions should adapt to the current *users' needs and usage patterns*. Indeed, since different users have different smartphone usage patterns, "it still remains a challenge how the goals [i.e., the interventions] needs to be appropriated to each user" (P03 [105], p. 16:23). Without taking into account the preferences of the user and their current needs, in particular, finding a trade-off between helping the user and avoiding unnecessary overloads becomes a challenge [214, 224]: as reported by Hiniker et al. [83] in their study on smartphone usage while caring for children, device-resistance is not universal, and some people may not have any desire or intention to limit their time with mobile devices. Finally, another kind of adaptability regards the *user* itself [112]. Besides using devices in different ways, indeed, users themselves have different levels of physical capabilities and technology competences. Ko et al. [112], in particular, highlight this challenge in the evaluation of *FamilyLink*, a parental-control DSCT that involves both parents and children.

5.1.2 Implementing Accurate Interventions. A challenge that is strictly related to adaptability is the need of implementing *accurate* interventions ($N = 20$, [11, 16, 33, 47, 80, 82, 101, 104, 106, 108, 133, 144, 159, 160, 162, 165, 189, 195, 223, 224]). According to the analyzed papers, researchers should improve the capabilities of DSCTs to detect the *contextual situations* in which the user is using their devices, as well as the current *usage intentions* of the user [33, 80, 82, 101, 104, 108, 133, 144, 160, 162, 165, 195, 223, 224]. People, indeed, often work at irregular hours [223] by using the same device both for work and leisure [108]. As highlighted by Oduor et al. [165], in particular, "a key design challenge is that 'work time' and 'home time' are not necessarily distinct" (p. 1324). At the same time, the same applications and websites with which users interact daily have often a "neutral" nature that is difficult to classify [33, 104, 165]. YouTube, for instance, can be used either for entertainment or for educational purposes. As a result, exactly defining what the user is currently doing with their devices [160] and what are their current feelings [195] are challenges that need to be solved in order to deliver the right intervention at the right time, e.g., to avoid "novelty effects" [224]. This is also reflected on the need of avoiding *bugs* [47, 106, 159, 189]. For example, an intervention delivered at an inappropriate time, e.g., a lockout task for a "good" or "meaningful" use [106], incurs considerable inconvenience to users. The same happens for bugs related to erroneous data tracking: wrong statistics affect the accuracy of the DSCT, which in turns affects usability [159]. While detecting such kind of bugs may be difficult, it is also an urgent challenge to increase the user's trust in the tool [47], thus reducing the risk of tool abandonment [189]. Finally, implementing an accurate DSCT also means allowing users to *understand* the visualized statistics and the delivered interventions [11, 16, 47, 108, 189]. Such a challenge mainly emerges from the complexity of the current visualizations adopted by DSCTs, which may result difficult to understand by novice users who are not used to interpreting charts [16]. Moreover, device usage statistics are often used in a retrospective manner, with users that may experience difficulties in remembering the exact purpose of their application usage [108]. Consequently, without accompanying statistics with sufficient details, e.g., contextual information, users cannot gain "a true understanding of their work patters from the data [47]" (p. 376), and they might not understand why an intervention has been delivered.

To make judgements about productivity and overuse, users therefore need "the details of what they were doing with their devices [189]" (p. 292).

5.1.3 Overcoming Technical Constraints. Researchers agree that a challenge to design and implement adaptable and accurate DSCTs is to overcome *technical constraints* ($N = 18$, [62, 80, 88, 101, 103, 104, 108, 115, 145, 149, 159, 160, 167, 189, 198, 214, 223, 224]). By investigating how to reduce the environmental impact of Internet through moderate use, for instance, Hill et al. [80] highlight that manipulating Internet traffic is not straightforward due to the current restrictions imposed by contemporary operating systems. More specifically, mobile operating systems have several restraints to preserve energy [224]. There are also differences between the different mobile operating systems available on today's market. The majority of DSCTs targeting mobile devices, indeed, are developed for Android-based smartphones [144], with different papers in our *Implemented Tools* corpus [62, 103, 149, 159] that highlight difficulties in developing the iOS counterpart. The Apple operating system, indeed, is typically more restrictive than Android, since it does not allow developers to implement fundamental operations that characterize the development of DSCTs, e.g., accessing sensitive information such as phone usage statistics [159], blocking the usage of other apps [103], muting notifications [103], launching background operations [103], and tracking Internet usage [80]. Moreover, even the operating system of Google is characterized by different versions that are continuously evolving[9], with each version that introduces new features and/or restrictions. DSCTs are particularly vulnerable to regular updates [115] and, at the same time, to software aging [198], i.e., the natural degradation of software capabilities. As these tools typically run on the same platforms that they aim to limit [198], indeed, changes in the hosting platforms may impact their functionality (see also the "hostile designs," through which the systems under modification become intentionally inhospitable [2, 59]). Without the possibility of developing stable and *cross-platform* solutions, these problems influence the deployment and the evaluation of DSCTs [62, 189]. Furthermore, technical constraints also hinder the possibility of developing *multi-device* DSCTs [88, 145, 160, 223]. As already reported in Section 3.1, targeting single devices is not sufficient to capture all the nuances of people's digital wellbeing [125], since many people work on their PC but spend large parts of their day online using phones and tablets [223].

5.1.4 Reducing Attrition. Minimizing attrition is another fundamental challenge that is mentioned in the analyzed papers ($N = 6$, [47, 82, 117, 118, 171, 198]). Attrition can be defined as the tendency of users to abandon a given technology after having using it for some time. DSCTs are particularly affected by this problem, since they "may over-restrict users in attempts to compensate for a lack of human accountability" (Schwartz et al. [198], p. 1), and, consequently, they may be abandoned by users because their settings do not match users' expectations or are too aggressive [118, 198]. Therefore, motivating the user to keep using a DSCT over long period of time is fundamental: according to numerous behavioral theories [178], indeed, the key aspect to change a behavior and forming a new habit is *repetition*, especially at the beginning of the process when the new behavior is starting to emerge. In the context of DSCTs, motivation is also important since using devices like smartphones is a socially situated practice, and it is therefore hard for an individual "to change a behavior when surrounded by social cues that run counter to his or her goals [82]" (p. 4755). Furthermore, as described by Collins et al. [47], motivating users is necessary to avoid an "acceptance epiphany," i.e., when users distractedly use the statistics displayed by the DSCT to accept their current behavior rather than investing effort to change it. Some papers on our *Implemented*

[9]The current version of Android, as of July 2021, is Android 11, but the majority of existing smartphones run a older version of the operating system, from Android 4.0 to Android 10 (see https://developer.android.com/about/dashboards for updated statistics).

Tools corpus explicitly try to address the attrition challenge. Schwartz et al. [198] outline "a new direction of research that is specifically focused on reducing risk of user abandonment of a given digital self-control tool" (p. 1). Their paper provides, in particular, a theoretical contribution of two principles that relate risk and efectiveness in DSCTs. Similarly, Kovacs et al. investigate the influence of periodically changing the intervention [117] and/or its difficulty [118] on users' attrition rate, while the *Aiki* tool of Inie and Lungu [88] has been designed to avoid adverse effects of blocking potentially necessary breaks.

5.1.5 Conducting Long-Term, Accurate Evaluations. Another challenge that is mentioned in the analyzed papers is related to the evaluation of DSCTs (N = 12, [33, 81, 116–118, 140, 145, 166, 167, 171, 198, 224]). Researchers, in particular, are aware that the effectiveness of DSCTs should be measured through *long-term evaluations* [81, 116, 117, 171, 198], preferably with follow-up periods [33, 171], since behavior change is a long and complex process: new habits, for instance, may take time to become established [178], and experiments [121] demonstrate that a new behavior needs from a few weeks to almost an year of repetition to become automatic, with substantial variation at individual level. While researchers are starting to recognize the need of evaluations that investigate the effectiveness of DSCTs in the long term, Section 6.1 shows that such a challenge is still underexplored. This may cast doubts on the promising results that typically accompany the studies included in our corpus (see Section 6.2). According to Kovacs et al. [116], in particular, while behavior change systems are typically effective during experiments, more effort should be put into evaluating "whether behavior change systems remain effective outside studies and bring longitudinal behavioral change" (p. 3). Besides long-term evaluations, researchers also point to the need of adopting *better measures* [33, 118, 140, 145, 167]. According to Lyngs et al. [145], for instance, most studies about Facebook overuse rely on self-reported measures, but prior work [57, 169] already demonstrated that self-report often does not correspond with actual use of digital devices. More explicitly, Okeke et al. [167] state that "self-reports capture what people say they do but not what they actually do, which is a major challenge in behavior change" (p. 1). Furthermore, the log of simple data like time spent or visits may not be sufficient to measure the effectiveness of DSCTs [224], but, at the same time, extracting more complex measures, e.g., users's sense of meaningfulness, is undoubtedly harder [140]. Finally, 2 papers in our corpus [145, 166] highlight that a factor that influence the evaluation of DSCTs is the fact that these tools typically combine a variety of different interventions at the same time, making it difficult to interpret causality [145] and understand the specific effects of a particular behavioral technique.

5.1.6 Dealing With The Contemporary Business Model. The last challenge that emerges from our analysis concerns the business model of contemporary tech companies (N = 8, [19, 139, 140, 144, 160, 204, 224]). In the "attention economy [54]", indeed, tech companies compete to grab users' attention and maximize advertisement revenue, by designing technologies that lead users into continuous and frequent usage. While Section 4.2 shows that researchers already investigated possible solutions to mitigate this one-sided view, in which technology is primarily designed to satisfy companies' revenue, dealing with (and possibly change) this business model is undoubtedly one of the most critical challenge that characterizes the field of interventions for digital self-control. DSCTs, indeed, may hinder business profitability. Widdicks and Pargman [224], in particular, pose the following open question: "how do companies create Internet services that encourage users to 'do business' with them, yet promote moderate Internet use and ensure users' do not become 'hooked' on their service?" (p. 6). There is therefore the need of exploring what could potentially motivate key stakeholders to support DSCTs [139]. While *designing for meaningfulness* may result in a lower user engagement in the short term, for example, it could also increase user loyalty in the long term [140]. All in all, designing technology that is also able to disincentive itself is, at the same time, "an open challenge and an ethical responsibility that HCI researcher should explore to counter problems like technology unwanted (over)use" (Monge Roffarello and De Russis [160], (p. 12).

5.2 Ethical Considerations in Papers on Digital-Self Control

This section investigates how ethics is taken into account in papers targeting digital self-control tools. We distinguish between a) researchers' ethics in conducting studies involving human subjects (Section 5.2.1), and b) ethical considerations for designing DSCTs (Section 5.2.2). To further investigate these two aspects, we applied the framework of Beauchamp and Childress [27] to understand whether and how the analyzed papers relate to core ethical principles of *autonomy*, *justice*, *non-maleficence*, and *beneficence*. As reported in Figure 5, researchers referenced the principle of autonomy in describing how they conducted their studies, while the principles of justice, non-maleficence, and beneficence were referenced by researchers when describing the proposed interventions.

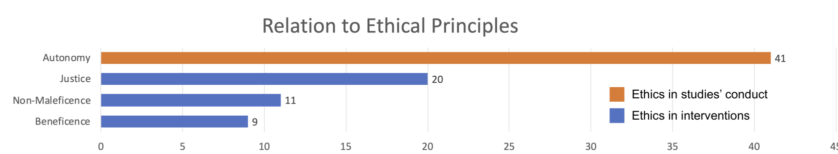

Fig. 5. The number of papers in which we found ethical principles of autonomy, justice, non-maleficence, and beneficence. We found the principle of autonomy in the description of the conducted user studies, and principle of justice, and non-maleficence, and beneficence in the proposed interventions. The principles are described by a state-of-the-art framework informed by healthcare ethics [27].

5.2.1 Ethics in Studies' Conduct. By analyzing all the research works in our *General Corpus*, we found that only 12 papers out of the 51 papers describing studies conducted with human subjects (23.52%, [64, 83, 139, 140, 161, 162, 165, 171, 173, 189, 198, 204]) mention a report of ethical approval for the conducted studies, e.g., through a revision of an Institutional Review Board (IRB). It is worth noting that most universities require IRB protocols as a prerequisite to human subjects works, and that researchers may use an IRB without explicitly mentioning it in their publications. Given the sensitive nature of topics like digital wellbeing and DSCTs, however, we believe that this kind of information should be always included.

Autonomy in Studies on Digital Self-Control. According to Beauchamp and Childress [27], the principle of *autonomy* refers to the respect for people's decision-making ability. A large amount of papers in our corpus ($N = 41$, [33, 47, 62, 65, 80, 81, 88, 103–106, 108, 114, 116, 117, 127, 133, 139, 140, 144, 145, 150, 159–162, 165–167, 167, 171–173, 186, 195, 198, 205, 213, 214, 223, 229]) address autonomy in describing the conducted studies from different perspectives. Overall, 15 papers [33, 116, 117, 127, 133, 149, 150, 161, 162, 171, 173, 186, 198, 213, 214] mention the usage of an *informed consent* to be accepted by users before starting the collection of their usage data. Such informed consents are also used to describe the study protocol [116], e.g., which data are collected, and to assure users that "their data would be kept anonymous" (P13 [149], p. 4). Other 15 papers take into account users' autonomy through the application or the discussion of *privacy* protecting measures [33, 80, 112, 114, 117, 145, 159, 165–167, 171, 186, 189, 198, 223], e.g., data *anonymization* [33, 145, 149, 159, 186, 214, 223], with the aim of "respecting and ensuring confidential treatment of peoples' personal information" (Thieme et al. [210], p.34:26). In the *Evaluated Tools* corpus, for instance, we found that Riedl et al. [186] obfuscate usage data that are related to websites that have not been explicitly added on lists of blocked sites by the user. Similarly, the browser extension developed by Lyngs et al. [145] collects data by using anonymous identifiers, and it does not store any "identifying information about the actual content engaged with" (p. 4). Rather than anonymizing data, Whittaker et al. [223] allow participants to review their collected logfiles before submitting them, with the possibility of removing entries that they do not wish to share. Similarly, Borghouts et al. [33] allow participants to delete or adapt "any sensitive or confidential information in their data, such as application and website names" (p. 32:17). NUGU by Ko et al. [114], instead, shares data by aggregating usage statistics among groups of people. Finally, 19 papers [47, 62, 81, 88, 104–106, 108, 114, 145, 160, 165, 171–173, 189, 195, 205, 229] describe the adoption of some inclusion criteria to recruit participants in their studies. Kim et al. [105], for example, evaluated their *GoalKeeper* DSCT by recruiting users who "considered themselves as excessive smartphone users and who were willing to reduce their use" (p. 16:10). Similar criteria are described in several other papers included in the *Evaluated Tools* corpus [47, 62, 81, 104, 106, 108, 114, 171–173, 205, 229].

5.2.2 Ethics in Interventions. Only 6 papers out of the 62 included in the *General Corpus* (0.08%) *explicitly* discuss or propose ethical guidelines related to the field of interventions for digital self-control., e.g., to inform the design of the implemented DSCTs and avoid problems like deception and lack of transparency [198]. The majority of these papers (N=4, [64, 80, 165, 224]), in particular, are included in the *Qualitative Discussions* corpus. In proposing hypothetical interventions to improve smartphone usage in social contexts, Genç and Coskun [64] acknowledge that the envisioned solutions may not be welcomed by users, as they may create unintended outcomes or violating users' privacy. Consequently, the authors highlight the need of "learning more about users' reactions towards the solutions generated to identify which of these is the most suitable ones for a specific context and a specific user group" (p. 10). As users perceive privacy as an important aspect to be considered when using DSCTs [112, 117, 159, 189], Okeke et al. [166] consider privacy *while* developing the proposed interventions. Their *Good Vibrations* tool, in particular, uses low-priority notifications that are not visible when the phone screen is locked, thus protecting users' privacy from prying eyes. Also Oduor et al. [165] elaborate on possible privacy issues that could derive from their interventions for limiting smartphone use at home, while Hill et al. [80] and Widdicks and Pargman [224] state that tracking and manipulating Internet access, e.g., to limit its usage, requires new policies and regulation, e.g., through additions to the European General Data Protection Regulation (GDPR). Given the importance of reducing Internet usage for environmental reasons, in particular, Widdicks and Pargman [224] envision an idealised world in which service providers "design applications and services that help users limit their Internet use [...], and create responsible, ethical and sustainable designs — utilising these aspects as a deliberative selling point to keep customers and profit" (p. 4). The same authors highlight the role of new policies as a mean to convince businessess that may be reluctant to introduce moderate Internet use within the design of their technologies. As some businesses may be reluctant to introduce moderate Internet use within the design of their technologies, researchers could engage with policy.

Justice in Digital Self-Control Tools. The ethical principle of justice focuses on the fair distribution of benefits, risks, and costs [191]. In the context of interventions for digital self-control, the papers that include such a principle ($N = 20$, [19, 81–83, 103, 108, 112, 113, 133, 139, 140, 160, 165, 166, 171, 189, 205, 213, 224, 229]) discuss the need of developing/evaluating DSCTs for/with wide and diverse populations. The majority of these papers [81, 83, 103, 108, 112, 113, 133, 139, 140, 160, 165, 171, 205, 213, 224, 229] cite such a need as a limitation or a future work to be explored, since the conducted studies often involve participants from the same geographical area [160, 213], e.g., Korea [113] or the Seattle metropolitan area [83], and with similar characteristics, e.g., students [112, 171] and workers [108]. This limits the generalizability of the retrieved findings [103, 108, 113, 140], which require additional research to be consolidated. Patterns of phone use, indeed, are likely to differ between different communities of users [83]. For this reason, in discussing the evaluation of their *FamilyLink* tool, Ko et al. [112] state that “additional research in different schools and cultural environments comprising students and parents from various socio-economic backgrounds is required” (p. 876). Liu et al. [133], instead, argue that designers should consider gender as a factor in persuasive design, since it “is important to match individual differences to maximize the motivation effect” (p. 775). Only 3 papers [19, 82, 166] describe an effort to include different populations in their studies. Hiniker et al. [82] state that their *MyTime* tool was evaluated with participants representing 20 different states and 4 races and ethnicities. Okeke et al. [166] describe an experiment that was conducted “with a diverse pool of participants, which is an advantage when compared to traditional non-online behavior change experiments that have been overwhelmingly conducted with participants from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies” (p. 4:9). Finally, the study of Aranda and Baig [19] includes participants coming from different countries, from Swiss to Argentina.

Beneficence and Non-Maleficence in Digital Self-Control Tools. The underlying motivation of all the research works included in the papers under analysis is, obviously, to positively contribute to users’ digital wellbeing. The principle of *beneficence*, however, is not only related to “doing good,” but also to “do it well” [23], i.e., to balance benefits against risks and costs with the aim of preventing harm. In this respect, only 9 papers in our corpus [64, 65, 105, 159, 160, 165, 198, 204, 224] explicitly declare to pursue a beneficence principle. For what concerns the implemented DSCTs (*Implemented Tools* corpus), for example, the *Time Sidekick* prototype [198] has been designed to meet principles of voluntariness, privacy, and non-deception, while *GoalKeeper* [105] focuses on the formation of “positive-long term goals related to various application domains such as entertainment, enlightenment, and sociality” (p. 16:5). Similarly, Monge Roffarello and De Russis [159] argue that DSCTs should promote the formation of new “healthy” habits, e.g., by increasing the usage of positive reinforcement techniques. Papers in the *Qualitative Discussions* corpus, instead, suggest researchers to adopt beneficence principles before designing interventions, e.g., to positively affect users [224] and to“design from a positive and hopeful perspective on engagement” (Genova et al. [65], p. 6).

Dually, the principle of *non-maleficence* can be seen as the explicit intention of not causing harm through the adopted strategies and in the conducted evaluations. Only a small subset of papers (N = 11, [64, 80, 88, 103, 108, 112, 113, 133, 165, 173, 189]) explicitly include such a principle. As acknowledged by some researchers in our *Implemented Tools* corpus [103, 133, 173], interventions adopted by DSCTs may have a negative impact on users’ wellbeing, thus motivating the need of carefully considering a non-maleficence principle in the design process. This means supporting more abstract representations that mitigate the privacy problem [112], reducing the frequency of feedback exposure [108], and avoiding too invasive and distractive reminders [113]. According to the researchers, in particular, using “punishment” methods has a double-bladed effect, since “it has higher effectiveness but also creates higher pressure on users” (P32 [133], p. 774). In discussing their virtual reality therapy for online-gaming addiction, Park et al. [173] argue that there could be ethical concerns surrounding the use of aversive stimuli, by highlighting the need to discuss the intensity of these interventions, since they could create “aversive stimuli that have the potential of provoking anxiety and defensive aggression" (p. 106). Kim et al. [103], instead, state that strict interventions like locking smartphones may infringe users’ autonomy, i.e., the state of being independent or self-governing, and argue that a direct social competition may result in a negative outcome. To mitigate these problems, the authors allow the users of their *Let’s FOCUS* tool, i.e., a DSCT with which students can share a "virtual room" with specific usage blocks, to leave a room at any time, and to self-organize social support groups. Furthermore, the authors abstract limiting records rather than displaying exact limiting records. As reported by some studies in the *Evaluated Tools* corpus [88, 189], the principle of non-maleficence can also be pursued while recruiting participants [88], as well as while collecting and analyzing data [189]. To avoid harm, in particular, the recruiting process of the study of Park et al. [173] involves the presence of a trained psychiatrist that screened participants through a structured clinical interview. Regarding data collection and analysis, instead, only the work of Rooksby et al. [189] report not having logged a device for ethical reason, since the user was using it while working with data about vulnerable people. Also some papers in the *Qualitative Discussions* corpus [64, 80, 165] provide concrete suggestions to take into account the non-maleficence principle in the digital wellbeing context, from taking into account possible unintended outcomes of DSCTs [64] to avoiding privacy-critical designs [165].

6 EVALUATIONS OF DIGITAL SELF-CONTROL TOOLS

This section analyzes the papers included in the *Evaluated Tools* corpus to investigate how researchers evaluate DSCTs, and to assess whether DSCTs are effective and can positively impact people's digital wellbeing (**RQ4**). We first analyze the procedures and methodologies described in the papers under analysis, by summarizing the recruiting process and involved participants, the collected measures (including the exploited tools and scales), and the adopted study designs (Section 6.1). Then, we report on the results of a meta-analysis that quantitatively investigates the effectiveness of the evaluated DSCTs (Section 6.2).

6.1 Evaluations of Digital Self-Control Tools

Table 8 provides a summary of all the studies assessing an implemented DSCT included in the 37 papers of the *Evaluated Tools* corpus. For each study, the table also reports the duration, the collected measures, the number of participants, and their demographic information, if available. Overall, we found a total of 43 distinct studies in the analyzed papers. The majority of the analyzed papers ($N = 32$) report on a single study[10], 2 papers [33, 223] that includes 2 studies, and 3 papers [47, 117, 118] that include 3 studies.

6.1.1 Recruiting Processes and Participants. Figure 6 summarizes the processes followed by researchers to recruit participants for the 43 studies included in our analysis. The majority of the studies ($N = 36$) include the description of a single procedure to recruit participants, ranging from *campus-wide campaigns* to *recruiting on Amazon's Mechanical Turk*, while 7 other studies mention the adoption of 2 of these procedures at the same time.

Table 8. Datasheet excerpt of the evaluations reported in the *Evaluated Tools* corpus. The 37 papers under analysis include a total of 43 distinct studies. For each of them, we report a summary, the number of participants and their demographic information (if available), the duration, and the collected measures.

Reference	**Summary**	**Participants [#]**	**Demographic**	**Duration [days]**	**Measures**
Riedl et al. [186]	Small scale controlled field study. Within-subject design with 1 week of baseline and 2 weeks of intervention.	4	Students and office workers (avg. age: 26)	21	General usage statistics, general perception, attrition.
Kim et al. [105]	Controlled field study. Within-subject design with 1 week of baseline and 3 weeks of intervention. Interventions were counterbalanced and rotated between participants.	36	Students (avg. age: 21.70)	28	General usage statistics, general perception, attrition, self-efficacy, self-control, smartphone addiction, focus, time-spent.

[10] [149] and [150] describe the same study with a different level of detail. The same happens for [103] and [102].

Table 8 Continued

Reference	Summary	Participants [#]	Demographic	Duration [days]	Measures
Lyngs et al. [145]	Controlled field study. Between subject design with control group and 2 weeks of baseline, 2 weeks of interventions, and 2 weeks of withdrawal. Follow-up survey after 5 months.	58	Students (avg. age: 22.50)	42	General perception, self-esteem, FOMO, Annoyance, unintended use, overuse, Facebook addiction, time-spent, number of visits, visits lenght, online behaviors (e.g., scrolling).
Pinder et al. [177]	Lab study. Pre-test/post-test between-subject design with control group.	40	Students (avg. age: 26.90)	1	Smartphone addiction, reaction time.
Kim et al. [106]	Controlled field study. Within-subject design with 1 week of baseline and 2 weeks of intervention.	40	Students (avg. age: 23.00)	21	General usage statistics, general perception, mental demand, physical demand, temporal demand, perfomance, effort, frustration, completion time, success rate, discouraged rate, time spent, number of visits.
Tseng et al. [214]	Controlled field study. Within-subject design with 1 week of baseline and 2 weeks of intervention.	15	Students, researchers, and office workers (avg. age: 34.50)	21	General usage statistics, general perception, productivity, focus, stress, self-control, time spent, number of visits.
Monge Roffarello et al. [159]	Controlled field study. Within-subject design with 1 week of baseline and 2 weeks of intervention.	38	Students and office workers (avg. age: 22.50)	21	General usage, statistics, general perception, smartphone addiction, self-regulation, time spent, number of visits.
Kovacs et al. [116]	Controlled field study. Within-subject design with 132 days of intervention. The frequency of the interventions varied across participants' groups.	2654	Not available (existing users of the tool)	132	Time-spent.
Kovacs et al. [117]	Controlled field study. Within-subject design with 3 weeks of intervention. The intervention strategy was rotated across participants.	217	Not available (existing users of the tool)	21	Time-spent, attrition;
Kovacs et al. [117]	Controlled field study. Between-subject design with 10 weeks of intervention. The number of active interventions varied across participants' groups.	409	Not available (existing users of the tool)	35	Attrition.
Kovacs et al. [117]	Controlled field study. Between-subject design. The design of the tool varied varied across participants' groups.	93	Not available (existing users of the tool)	10	Attrition.
Okeke et al. [166]	Controlled field study. Between-subject design with control group. The intervention strategy varied across participants' groups. 1 week of baseline, 1 week of intervention, and 1 week of withdrawal. Follow-up interview after 1 year.	50	Amazon's Mechanical Turk users (avg. age: 28.80)	21	General perception, time-spent, number of visits.
Mark et al. [149]; Mark et al. [150]	Controlled field study. Within-subject design with 1 work-week of baseline and 1 work-week of intervention.	31	Office workers	10	General usage statistics, general perception, focus, productivity, workload, susceptibility to distraction.
Kim et al. [104]	Controlled field study. Between-subject design with control group. The tool was compared with a state-of-the-art intervention strategy. 1 week of baseline and 2 weeks of intervention.	40	Students (avg. age: 26.50)	21	General usage statistics, general perception, attrition.
Kim et al. [103]; Kim et al. [102]	Field deployment. Within-subject design.	379	Students	41	General usage statistics, general perception, usability.
Foulonneau et al. [62]	Controlled field study. Between-subject design with 1 week of baseline and 5 weeks of interventions. The intervention strategy varied across participants' groups.	19	Anonymous users (avg. age: 32.50)	42	Time-spent.
Morley et al. [161]	Lab study. Within-subject design.	23	Students	1	Realism, anxiety.
Ko et al. [113]	Field deployment. Within-subject design.	976	Students (avg. age: 22.59)	25	General usage statistics, general perception, distraction score.
Kim et al. [108]	Controlled field study. Between-subject design with 2 weeks of baseline, 4 weeks of intervention, and 2 weeks of withdrawal. The intervention strategy varied across participants' groups.	24	Students and office workers, (avg. age: 27.88)	56	General usage statistics, general perception, productive rate, attrition, self-awareness, self-reflection.

Table 8 Continued

Reference	Summary	Participants [#]	Demographic	Duration [days]	Measures
Hiniker et al. [82]	Controlled field study. Within-subject design with 1 week of baseline and 1 week of intervention.	23	Anonymous users (avg. age: 33.50)	14	General usage statistics, General perception, time-spent.
Whittaker et al. [223]	Controlled field study. Within-subject design with 2 days of baseline and 2 days of intervention.	61	Students and office workers (avg. age: 29.40)	4	General perception, general usage, time-spent, number of visits.
Whittaker et al. [223]	Controlled field study. Between-subject design with 2 days of baseline and 2 days of intervention. The intervention strategy varied across participants' groups.	57	Students (avg. age: 20.80)	4	Focus, university grades, adopted strategies.
Rooksby et al. [189]	Field deployment. Within-subject design.	21	Students (avg. age: 23.05)	At least 28 days	General usage statistics, general perception.
Ko et al. [112]	Controlled field study. Within-subject design with 1 week of baseline and 2 weeks of intervention.	35	Parents and children (avg.age: 32.02)	21	General usage statistics, general perception, time-spent, number of visits, permissive style, authoritarian style, authoritative style.
Collins et al. [47]	Controlled field study of a commercial tool. Between-subject design. The intervention strategy varied across participants' groups.	16	Students (avg. age: 23.75)	10	General usage statistics, time-management skills, stress, time-spent.
Collins et al. [47]	Controlled field study of a commercial tool. Between-subject design. The intervention strategy varied across participants' groups.	30	Students (avg. age: 20.33)	10	General usage statistics, time-management skills, stress, time-spent.
Collins et al. [47]	Field deployment of a commercial tool. Within-subject design.	7	Students, researchers, and office workers (avg. age: 32.5)	At least 14 days	Utility, attrition.
Ko et al. [114]	Controlled field study. Between-subject design with 1 week of baseline and 2 weeks of intervention.	62	Students (avg. age: 25.74)	21	General usage statistics, general perception, time-spent, number of visits, smartphone addiction, attrition, self-efficacy.
Löchtefeld et al. [134]	Field deployment. Within-subject design.	11700	Not available (existing users of the tool)	February to October, 2013	General usage statistics, general perception.
Liu et al. [133]	Lab study. Between-subject design with control group. The intervention strategy varied across participants' groups.	30	Students	1	Off-task time, voluntary switching, stress.
Park et al. [173]	Prospective trial. Between-subject design with control group. Pre-treatment evaluation, active treatment, and follow-up evaluation. The treatment was compared with a state-of-the-art intervention strategy.	36	Students (avg. age: 23.70)	28	Internet addiction, depression, anxiety, ADHD, amplitude of low-frequency fluctuation, functional connectivity, posterior cingulate cortex.
Ahn et al. [127]	Field deployment. Within-subject design.	14	Anonymous users (avg. age: 26.57)	At least 7 days of usage	Smartphone addiction, time-spent, number of visits.
Park et al. [172]	Field deployment. Within-subject design.	7	Students (avg. age: 21.57)	4	General usage statistics, general perception, attrition, self-reflection.
Borghouts et al. et al. [33]	Field deployment. Within-subject design with 1 week of baseline and 1 week of intervention.	9	Office workers	14	General usage statistics, general perception, off-task time.
Borghouts et al. et al. [33]	Online experiment. Between-subject design with control group. The intervention strategy varied across participants' groups.	47	Students, office workers (avg. age: 29.30)	1	Off-task time,completion time.
Hiniker et al. [81]	Controlled field study. Within-subject design with 1 week for each condition.	24	Parents and children	21	General usage statistics, general perception, time-spent,ambient audio.
Park et al. [171]	Controlled field study. Between-subject design with 1 week of baseline and 3 weeks of intervention. The intervention strategy varied across participants' groups.	210	Students (avg. age: 24.07)	28	General usage statistics, general perception, time-spent, success rate.

Table 8 Continued

Reference	Summary	Participants [#]	Demographic	Duration [days]	Measures
Song et al. [205]	Controlled field study. Within-subject design with 1 week of baseline, 1 week of intervention, and 1 week of withdrawal.	3	Students, office workers	21	General usage statistics, general perception.
Zhou et al. [229]	Field deployment. Within-subject design.	26	Not available (existing users of the tool)	244	General usage statistics, general perception, time-spent.
Inie et al. [88]	Field deployment. Within-subject design.	10	Office workers	14	General usage statistics, general perception, success rate.
Kovacs et al. [118]	Field deployment. Within-subject design.	1240	Not available (existing users of the tool)	Variable (200 usages per user)	Preferences on difficulty level.
Kovacs et al. [118]	Controlled field study. Between-subject design. The frequency of the interventions varied across participants' groups.	1108	Not available (existing users of the tool)	528	Attrition.
Kovacs et al. [118]	Field deployment. Within-subject design.	644	Not available (existing users of the tool)	385	Preferences on difficulty level.

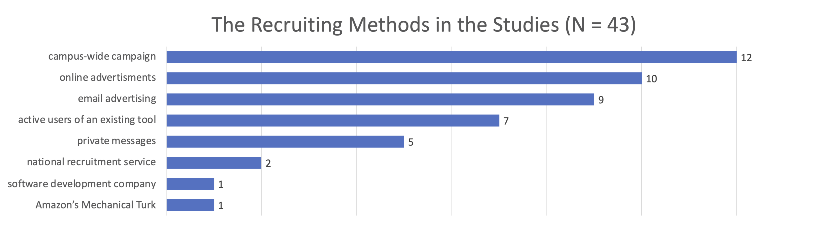

Fig. 6. The recruiting methods reported in the studies included in *Evaluated Tools* corpus. Researchers mainly recruited participants through posters and announcements in their universities, or by means of mailing lists and online advertisements.

As Figure 6 shows, the most common recruiting method in DSCTs studies (N = 12, [102, 103, 106, 108, 113, 145, 161, 173], 2 studies in [223], and 2 studies in [223]) is a *campus-wide campaign*, e.g., by means of promotional posters and announcements in the researchers' universities [47, 103, 113, 161]. Common strategies also include online posts (N = 10, [33, 47, 88, 104, 105, 108, 114, 145, 171, 229]), e.g., on social networks [108, 145], the usage of mailing lists (N = 9, [88, 108, 112, 145, 149, 150, 159] and 2 studies in [33]), private messages to the researchers' social circles (N = 4, [62, 127, 133, 159, 205]), and the recruitment of users that were already using an existing DSCT (N = 7, [116] and 3 studies included in [117] and [118], respectively).

Overall, the median number of recruited participants is 36, with a minimum of 4 involved users [186] and a maximum of 11,700 participants of the study of Löchtefeld et al. [134], who analyze the data of all the users who installed the *AppDetox* app from February to October, 2013. By analyzing the 26 studies that report age information, we can see that the users involved in DSCTs studies are on average 26.23 years old (SD = 4.14). As shown in Figure 7, they are most often students (N = 25, [33, 103–106, 108, 113, 114, 133, 145, 159, 161, 171–173, 177, 186, 189, 205, 214], the 3 studies in [47], and the 2 studies in [223]). Together with the most common recruiting method reported above, i.e., campus-wide campaigns, this highlights a strong selection bias towards young university students. Another recurrent occupation for participants in DSCTs studies is *office workers*

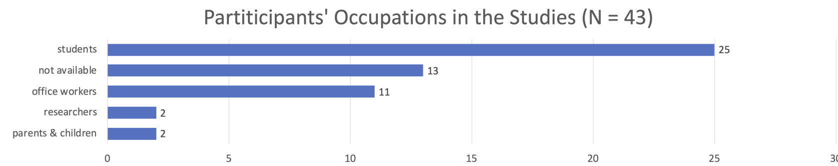

Fig. 7. The participants' occupations reported in the studies included the *Evaluated Tools* corpus. As reported in Table 8, 13 studies do not report any participants' occupation, e.g., because researchers analyze data of existing users of the tool [116–118, 134]. All in all, the figure shows that there is a strong bias towards university students and office workers.

($N = 11$, [47, 88, 108, 149, 159, 186, 205, 214, 223] and 2 studies in [33]). Least common occupations are *researchers* ($N = 2$, [47, 214]) and a generic *"parents and children"* ($N = 2$, [81, 112]). Therefore, we can identify another selection bias. Our findings, indeed, highlight that DSCTs are nearly always evaluated with technology savvy users that use devices like PCs and laptops every day for studying and working.

6.1.2 Measures, Tools, and Scales. To evaluate a DSCT, researchers collect different measures related to the *tool* usage, its *influence* on usage patterns, as well as its influence on *users'* character traits and feelings. Figure 8 shows the measures, i.e., the logged and/or collected data, that are mentioned in at least 4 of the 43 studies included in our corpus.

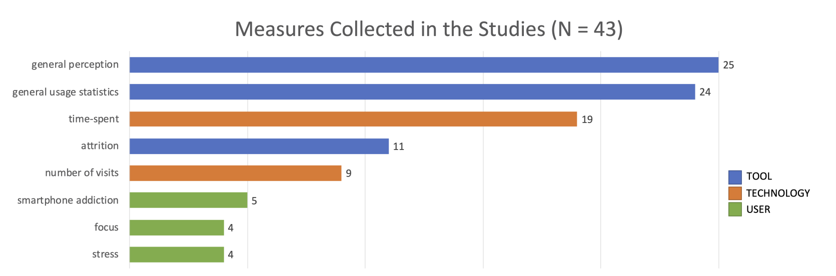

Fig. 8. Measures collected in at least 4 of the 43 studies included in our corpus. Collected measures focus on the usage of the tools under analysis (blue bars), on the influence of the tools on technology use (orange bars), and on users' character traits and feelings (green bars).

Tool Usage. For what concerns the measures related to the tool usage (the blue bars of Figure 8), the majority of the studies in our corpus ($N = 24$, [33, 81, 82, 88, 103–106, 108, 112–114, 134, 149, 159, 171, 172, 186, 189, 205, 214, 229] and 2 studies in [47]) include an assessment of the evaluated DSCTs through a generic analysis of their *usage statistics*, i.e., to understand whether and how participants used the tools during the studies. Such an assessment investigates different aspects of DSCTs usage, from analyzing how many times a user accessed the tool to consult her device usage statistics, to extracting information about how many times an intervention has been delivered and how many times it has been respected (or not) by the user. Collins et al. [47], for instance, measure the mean number of days *RescueTime* was accessed by participants to control their social networks use. Tseng et al. [214], instead, log all the participants' interactions with their *UpTime* tool, from editing the list of distractive websites to starting or cancelling a Pomodoro blocking session. Similarly, Monge Roffarello and De

Russis [159] analyze the number of smartphone usage timers and blockers that have been respected, snoozed, and deleted by participants, respectively. Stemming from these statistics, researchers also ask participants their *general perception* about the usage of the evaluated tools (N = 25, [33, 81, 82, 88, 103–106, 108, 112–114, 134, 145, 149, 159, 166, 171, 172, 186, 189, 205, 214, 223, 229]), e.g., by asking participants qualitative questions about their experience with the software [149] like "what did you think of the app [189]" (p. 288), or by soliciting short stories about how participants used the DSCT [113]. Another common collected measure that characterizes the usage of the evaluated tools (N = 11, [47, 104, 105, 114, 114, 118, 172, 186] and the 3 studies in [117]) is related to *attrition*, i.e., the the users' tendency to stop using the tool.

Influence on Usage Patterns. Regarding the influence of the evaluated DSCTs on technology use (the orange bars of Figure 8), the analyzed studies report on 2 main measures, i.e., *time-spent* (N = 19, [62, 81, 82, 105, 106, 112, 114, 116, 117, 127, 145, 159, 166, 171, 214, 223, 229] and 2 studies in [47]) and *number of visits* (N = 9, [106, 112, 114, 127, 145, 159, 166, 214, 223]). By measuring time-spent, researchers analyze the influence of the DSCT on the time the user is spending over a particular technological source that is subjected to an intervention. Researchers, in particular, analyze the global time spent by users on a device (see the first version of *Socialize* [159], which timers can affect the overall usage of the smartphone), and/or the time spent on the specific mobile apps (e.g., [82]) or websites (e.g., [117]) on which the user set up an intervention. Some studies even focus on (and measure) time-spent on specific services, only, e.g., Facebook [145]. Number of visits are strictly related to time-spent, and they count how many distinct times the user is accessing a device (e.g., [159]), a mobile app (e.g., [106]), or a website (e.g., [214]).

Influence on Users' Character Traits and Feelings. To measure the influence of the evaluated DSCTs on users' character traits and feelings (the green bars of Figure 8), researchers mainly focus on assessing the participants' self-perception of *smartphone addiction* (N = 5 [105, 114, 127, 159, 177]). Kim et al. [105] use such a metric to characterize participants in the initial meeting of their study [105], while Pinder et al. [177] use it as an independent variable to evaluate their different groups of participants. Monge Roffarello and De Russis [159] and Ko et al. [114], instead, collect smartphone addiction perception at the beginning and end of their studies, with the aim of investigating the influence of *Socialize* and *NUGU*, respectively, on the how participants perceive their problematic smartphone usage. In another study, Ahn et al. [127] elicit smartphone addiction from their participants to evaluate the reliability and efficacy of their *SAMS* system, that has been developed to automatically and objectively assess users' smartphone addiction. Besides smartphone addiction, some studies also measure the influence of the evaluated DSCT on users' *focus* (N = 4, [105, 149, 214, 223]) and *stress* (N = 4, [133, 214] and 2 studies in [47]), e.g., to understand whether and how the tool "support the daily work focus of the user [105]" (p. 16:3), or to investigate "whether different visualizations or persuasion strategies would cause more psychological effects such as stress [133]" (p. 770).

To collect the aforementioned measures, researchers adopt several instruments, i.e., state-of-the-art scales and evaluation tools like questionnaires and interviews. Table 9 reports all the instruments that are mentioned in the 37 papers of the *Evaluated Tools* corpus.

Table 9. The different scales and evaluation tools used in the 37 papers under analysis to evaluate the characteristics of the studies' participants as well as their reaction to the usage of the involved DSCTs.

Scale	Summary	Paper/Authors
Smartphone Addiction Scale (SAS) [120]	A self-diagnostic scale composed of 33 questions and 6 points from daily-life disturbance to tolerance to evaluate the smartphone addiction using self-reporting.	Kim et al. [105]; Pinder et al. [177]; Monge Roffarello et al. [159]; Ko et al. [114]; Ahn et al. [127];
Nasa-Task Load Index (NASA-TLX) [76]	A multidimensional assessment tool to evaluate perceived workload of a task or a system in terms of mental, physical, and temporal demand, performance, effort, and frustration.	Kim et al. [105]; Kim et al. [106]; Mark et al. [149]; Mark et al. [150]; Liu et al. [133];
Cognitive Absorption Scale [9]	A scale to measure five states associated with deep engagement with technology, i.e., temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity.	Kim et al. [105]; Mark et al. [149]; Mark et al. [150];
General Self-Efficacy Scale (GSE) [92]	A 10-item psychometric scale to assess optimistic self-beliefs and perceived self-efficacy to cope with daily activities and isolated stressful events.	Monge Roffarello et al. [159]; Ko et al. [114];
UPPS Impulsivity Scale [53, 222]	A 59-item self-report to assess five distinct dimensions of impulse behavior in adolescents and adults, i.e., urgency, premeditation, perseverance, sensation seeking, and positive urgency.	Mark et al. [149]; Mark et al. [150];
Passive and Active Facebook Use Measure (PAUM) [67]	A questionnaire to measure the frequency of activities on Facebook according to the usage dimensions of active social use, e.g., chatting, active non-social use, e.g., tagging, and passive use, e.g., browsing.	Lyngs et al. [145];
Usefulness, Satisfaction, and Ease of Use Questionnaire (USE) [141]	A 30-item survey questionnaire to measure the subjective usability of a product or service through four dimensions of usability, i.e., usefulness, ease of use, ease of learning, and satisfaction.	Kim et al. [103];
Parental Authority Questionnaire (PAQ) [37]	A 30-item questionnaire to measure permissive, authoritarian, and authoritative parental authority prototypes for both the mother and the father.	Ko et al. [112];
Time Management Behavior Scale (TMBS) [146]	A 33-items questionnaire to identify and describe factors related to time management, i.e., goals and priorities, mechanics, preference for organisation, and perceived control of time.	Collins et al. [47];
Young's Internet Addiction Scale (YIAS) [227]	A 20-item questionnaire to evaluate the severity of users' Internet addiction according to client motivation, online time management, improved social relationships, improved sexual functioning, engagement in offline activities, and ability to abstain from problematic applications.	Park et al. [173];
Multidimensional Facebook Intensity Scale (MFIS) [170]	A 13-item questionnaire to predict and measure Facebook-related activities like liking and posting according to four factors, i.e., persistence, boredom, overuse, and self-expression.	Lyngs et al. [145];
Single-Item Self-Esteem Scale (SISE) [187]	A single-item questionnaire ("I have high self-esteem") to measure global self-esteem.	Lyngs et al. [145];
Smartphone Addiction Proneness Scale (SAPS) [100]	A 15-item survey to asses the level of addiction towards smartphones according to disturbance of adaptive functions, virtual life orientation, withdrawal, and tolerance.	Ko et al. [113];
Parent-Adolescent Communication Scale (PACS) [22]	A 20-item questionnaire to measure communication between a parent and an adolescent with two sub-scales, i.e., openness and problems.	Ko et al. [112];
Perceived Stress Scale (PSS) [46]	A 14-item questionnaire to measure the degree to which situations in individual's life are appraised as stressful. It asks about feelings and thoughts during the last month.	Collins et al. [47];
Structured Clinical Interview for DSM-IV-TR Axis I Disorders Patient Edition (SCID-I/P) [60]	A diagnostic exam used to determine DSM-IV Axis I disorders (major mental disorders). SCID-I/P is designed to be administered by a mental health professional.	Park et al. [173];
Beck's Depression Inventory (BDI) [29]	A 21-item multiple-choice self-report inventory for measuring the severity of depression. It is composed of items ranging from cognition factors, e.g., guilt, to physical symptoms, e.g., fatigue and weight loss.	Park et al. [173];
Beck's Anxiety Inventory (BAI) [28]	A 21-item multiple-choice self-report inventory for measuring the severity of anxiety in children and adults. It asks about common symptoms of anxiety that the individual suffered in the last week, e.g., numbness and tingling.	Park et al. [173];
WHO ADHD Self-Report Scale (ASRS) [99]	A 18-item questionnaire about frequency of recent DSM-IV Criterion A symptoms of adult ADHD.	Park et al. [173];

The most common instruments, mentioned in at least 2 papers, are the following:

Smartphone Addiction Scale. The Smartphone Addiction Scale (SAS) [120] is a self-diagnostic scale of 33 questions with a six-point Likert scale. Its aim is to evaluate smartphone addiction using self-reporting according to 6 factors that can be influenced by smartphone usage: daily-life disturbance, positive anticipation, withdrawal, cyberspace-oriented relationship, overuse, and tolerance. Together with its short version for adolescents (SAS-SV [119]), the Smartphone Addiction Scale was originally developed and validated for use with South

Korean users. The popularity and the adoption of this scale has rapidly grown, and SAS is now validated for different countries, ranging from USA [73] to Italy [55] and Spain [138]. The Smartphone Addiction Scale is mentioned in 5 papers [105, 114, 127, 159, 177] in the *Evaluated Tools* corpus. In their studies located in Korea, Ko et al. [114] used the original SAS version, while Lee et al. [127] adopted K-SAS [180], a specialization of SAS for Korean adults. Other researchers [105, 159, 177], instead, use the short version of the scale for adolescents (SAS-SV).

NASA-Task Load Index. NASA-Task Load Index (NASA-TLX) [76] is an assessment tool to evaluate perceived workload of a task or a system. It incorporates a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales, i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration. It was originally developed by the Human Performance Group at NASA's Ames Research Center with more than 40 laboratory simulations, and it is now the standard de facto for measuring perceived workload, especially in human factors research [75]. Overall, 5 papers [105, 106, 133, 149, 150] in the *Evaluated Tools* corpus use the NASA-TLX tool to assess the workload needed by the user to use a DSCT, e.g., to complete a lockout task [106], or to measure how the user's workload is affected by the adopted DSCT [105, 133, 149, 150], e.g., during a task on a PC [149, 150].

Cognitive Absorption Scale. Cognitive absorption refers to a situation-specific and individual state of deep involvement with a technology [10], e.g., a software. According Agarwal et al. [9], the three fundamental elements that characterize cognitive absorption are the state of flow [50], i.e., the individual's experience of complete involvement in an activity, the trait of absorption [208], i.e., the individual's experience of being in a state of deep attention, and the notion of cognitive engagement [221], i.e., a state encompassing individual's attention focus, engagement, and interest. The same authors proposed the Cognitive Absorption Scale [9] to measure five states associated to cognitive absorption, i.e., temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity. Such a scale is mentioned by 3 papers [105, 149, 150] in the *Evaluated Tools* corpus, which measure cognitive absorption during daily [105] or working activities [149, 150].

General Self-Efficacy Scale. General self-efficacy can be defined as "the belief in one's competence to cope with a broad range of stressful or challenging demands" [142]. The General Self-Efficacy Scale (GSE) was originally developed by Jerusalem and Schwarzer [92] as a 10-item psychometric scale to assess optimistic self-beliefs and perceived self-efficacy to cope with daily activities and isolated stressful events. During the years, it has been revised and adapted to more than 25 languages, thus becoming a universal construct [196]. Overall, 2 papers [114, 159] in the *Evaluated Tools* corpus adopt the General Self-Efficacy Scale. Both Ko et al. [114] and Monge Roffarello and De Russis [159], in particular, customized the Korean version of the General Self-Efficacy Scale [200] to the context of self-regulation of smartphone use.

UPPS Impulsivity Scale. The UPPS Impulsivity Scale was originally developed by Whiteside and Lynam [222] to standardize which traits are measured in the different existing measures of impulsivity. The original scale assesses, through self-report, 4 traits of impulse behavior in adolescents and adults: *negative urgency,* i.e., the tendency to act rashly under extreme negative emotions, *lack of premeditation*, i.e., the tendency to act without thinking, *lack of perseverance*, i.e., the inability to remain focused on a task, and *sensation seeking*, i.e., the tendency to seek out novel and thrilling experiences. Stemming from the original scale, Cyders et al. [53] developed UPPS-P, a 59-item version that integrates *positive urgency*, i.e., the tendency to act rashly under extreme positive emotions. Also, the UPPS-P has been adapted to several languages, from Korean [131] to French [31]. The UPPS Impulsivity Scale is mentioned in 2 papers [149, 150] in the *Evaluated Tools* corpus. The scale, in particular, is used in the study of Mark et al. (described in 2 different papers [149, 150]) to "explore the role of individual differences when distractions are reduced [149]" (p. 3).

6.1.3 Implemented Designs and Studies Characteristics. The median duration of a study evaluating a DSCT is 21 days, with a minimum of 1 day and a maximum of 528 days. Only 5 studies have lasted more than two months: 2 experiments described in [118], and the studies in [116, 134, 229]. All these studies are based on existing DSCTs, be they research artifacts or commercially available solutions, with researchers that had the opportunity to exploit data coming from users who had already installed the tool in the past.

Figure 9 summarizes the designs and the main characteristics of the 43 studies included in the *Evaluated Tools* corpus. Overall, 26 studies (those included in [33, 47, 81, 82, 88, 103, 105, 106, 112, 113, 116, 117, 127, 134, 149, 159, 161, 172, 186, 189, 205, 214, 223, 229] and 2 experiments in [118]) follow a *within-subject design*, in which all participants can use all the functionality of the evaluated DSCTs, i.e., they are exposed to the same interventions. Such evaluations include *controlled field studies* [47, 81, 82, 105, 106, 112, 116–118, 149, 159, 186, 205, 214, 223], *field deployments* [33, 47, 88, 103, 113, 118, 127, 134, 172, 189, 229], and *lab studies* [161]. During within-subject controlled field studies, researchers maintain some degree of control over the DSCT. Mainly, this is related to the possibility of deciding when to activate the provided interventions [33, 81, 82, 105, 106, 112, 149, 159, 186, 205, 214, 223]. This allows researchers to collect *baseline* data, i.e., how users behave without the help of the DSCT, and *intervention* data, i.e., how users interact with their technological sources when using the DSCT. In some cases, researchers even control the kind of interventions that are delivered to the user (see the "rotating" interventions of *HabitLab* [117]), as well as the frequency of the interventions [47, 116], i.e., how many times on average an intervention like a persuasive message is delivered to the user. Through within-subject field deployments, instead, researchers recruit participants to install and freely use a DSCT for a given amount of time, without enforcing any baseline periods nor changes in the DSCT functioning. Such field deployments are used to test novel DSCTs [33, 88, 103, 113, 118, 127, 134, 172, 189], as well as to evaluate commercial DSCTs (see the third study carried out by Collins et al. [47], who interviewed *RescueTime* users, and the study of Zhou et al. [229], who evaluated an existing intervention in an online gaming platform). Only Morley et al. [161] report on a within-subject lab study. The authors present a design of a virtual reality user experience to evaluate users' smartphone and driving behaviors. During the 1-day lab study, participants interact with their smartphones while driving a virtual car, and they experience (virtual) crash events that should serve as a deterrent for smartphone use while driving.

Another set of papers include studies that follow a *between-subject design* ($N = 17$, [33, 62, 104, 108, 114, 118, 133, 145, 166, 171, 173, 177, 223], 2 studies in [117], and 2 studies in [47]). Differently from within-subject studies, between-subject evaluations have two or more groups of subjects each being tested by a different study condition. While within-subject designs typically require fewer participants and are cheaper to run, between-subjects reduce learning effects and transfer of knowledge across the different tested conditions. As reported in Table 8, between-subject evaluations include *controlled field studies* [33, 47, 62, 104, 108, 114, 117, 118, 145, 166, 171, 223], *lab studies* [133, 177], and *prospective trials* [173]. Researchers use controlled field studies that follow a between-subject design to test, in-the-wild, different variations of the implemented DSCTs. Such variations can be small, e.g., asking users to provide daily retrospective estimations at noon or midnight [47], or they may involve completely different intervention strategies (see Lyngs et al. [145], who tested two different intervention strategies against Facebook overuse, i.e., goal advancement and removing the newsfeed). Between-subject lab studies, instead, are used both to assess intervention variations [133] or to evaluate the effects of an intervention strategy with respect to a control group [177]. A between-subject design is also followed by Park et al. [173], who carried out a 4-week prospective trial to compare their virtual reality therapy for online gaming addiction with a state-of-the-art psychological treatment named cognitive behavior therapy [107].

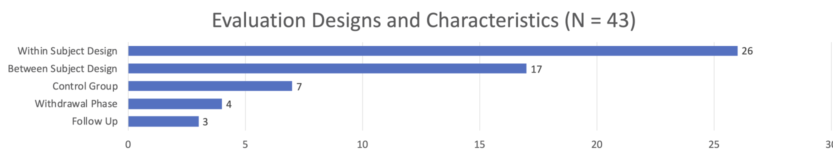

Fig. 9. A summary of the designs and the main characteristics of the 43 studies under analysis. Researchers evaluate their DSCTs through both within-subject and between-subject evaluations, but they rarely include a control group, and even more rarely they study the long-term effectiveness of their solutions through follow-ups and withdrawal phases.

All in all, Table 8 shows that the most common type of evaluation adopted by researchers is a 21-days controlled field study that follow a within-subject design [81, 106, 112, 159, 186, 205, 214]. During such an evaluation, the DSCT is deployed on participants' smartphones, PCs, or both, and usage data are typically collected for 1 initial week of baseline, i.e., 7 days during which the tool is "transparent" to the participants, and for the subsequent 2 weeks of intervention, i.e., 14 days during which participants can use all the functionality of the DSCT. By comparing usage statistics without and with the DSCT, researchers aim at analyzing the influence of the tool on technology usage. Despite common, we argue that such a design is not sufficient to accurately understand the effectiveness of DSCTs. The same researchers, indeed, agree on the need of longer-term evaluations, since behavior change is a long and complex process (see Section 5). Furthermore, despite the baseline phase, the lack of a "real" control group, i.e., a group of participants who do not receive the experimental treatment, may result in inaccurate evaluations due to problems like confirmation biases, i.e., the tendency for experimenters to give their expected outcome too much weight when measuring results [163]. As shown in Figure 9, the lack of a control group is also common among between-subject studies. Indeed, only 7 studies out of 35 (20%) compare participants using a DSCT with a group of people who do not use any interventions [33, 104, 133, 145, 166, 173, 177]. We can conclude that such a shortcoming profoundly characterizes how contemporary DSCTs are evaluated.

With a few notable exceptions, e.g., the evaluations of *HabitLab* [116] and *AppDetox* [134], Table 8 also demonstrate that short-term studies characterize the majority of the papers included in our corpus, independently of the adopted study design. A very limited number of researchers try to overcome such a problem through follow up interviews and surveys [145, 166, 173], e.g., 5 months after the end of the study [145]. Unfortunately, this does not provide a complete and accurate assessment of the DSCTs long-term effects, since self-report is typically different than actual use of digital devices [57, 169]. Moreover, only 4 studies [108, 145, 166, 205] include (and monitor) a *withdrawal* phase, i.e., a phase during which the delivered interventions are removed. Consequently, it is hard to understand whether the effects of a DSCT would survive a break in the use of the tool.

6.2 Meta-Analysis on DSCTs Effectiveness

We run a meta-analysis to quantitatively investigate the effectiveness of the evaluated DSCTs in reducing people's time using devices or applications (**RQ4**).

6.2.1 Measures and Procedure. We measured the effectiveness of a DSCT by analyzing its influence on the *time spent* on a target device, website, or mobile application. While time spent on a given online or mobile content does not perfectly correspond to attention or engagement behavior [149], e.g., as users can get distracted by "offline" events, prior work has generally accepted it as an effective estimation of users' engagement with technology [223]. Furthermore, time spent is one of the most common metrics collected during the studies evaluating DSCTs (see Section 6.1.2). For DSCTs targeting the overall usage of a device, e.g., the first version of Socialize [159], we analyzed the effectiveness of the tool in reducing the overall time spent by the user with the device. For DSCTs targeting specific websites or mobile applications, e.g., the Facebook interventions explored by Lyngs et al. [145], we analyzed the effect of the tool on the overall time spent on the specific service, e.g., Facebook.

Given this definition of effectiveness, we re-analyzed the *Evaluated Tools* corpus to select the papers to be included in our meta-analysis. We applied, in particular, the following additional criteria:

- papers that study the impact of the proposed DSCTs on a time-spent metric;
- papers that compare usage data between a "baseline phase," i.e., when there is no support from the DSCT, and an "intervention phase," i.e., when the DSCT supports the user[11].

Furthermore, we excluded papers that did not report sufficient statistical data, e.g., those reporting means without standard deviations. From the initial corpus, we selected 9 papers as eligible to be included in the meta-analysis, each of which included exactly one study investigating whether the evaluated DSCT reduced the time spent by the user on a given device, web site, or mobile application. Since only a small portion of these studies reported an explicit effect size, we decided to run the meta-analysis by leveraging means, standard deviations, and sample sizes. For each of the 9 studies, in particular, we extracted these information from both the "baseline phase" and the "intervention phase." Overall, 6 studies (S1-6) [112, 114, 159, 171, 214, 223] explicitly reported all the needed information. For two studies (S7-8) [82, 105], instead, we estimated means and standard deviations from the reported figures, while for another study (S9) [145] we calculated these information from the provided dataset. When multiple studies were available, e.g., when a paper compared different configurations of the same interventions, we included data for the best-performing intervention strategy, only.

To run the meta-analysis, we exploited *R* [183], a language and environment for statistical computing. We followed, in particular, the guide *"Doing Meta-Analysis in R"* by Harrer et al. [72]. Since the analyzed studies involved different types of users, we applied a Random-Effects-Model [32], by estimating the variance of the distribution of true effect sizes (τ^2) through the Hartung-Knapp-Sidik-Jonkman (HKSJ) method [89]. We used Hedges' g as the primary estimate of effect size for each DSCT. Hedges' g is defined as the difference between the two means (for baseline and intervention phases, respectively) divided by the pooled standard deviation. We used the Cohen's guidelines [45] to interpret effect sizes. According to Cohen, $g = 0.2$ suggests a small effect, $g = 0.5$ suggests a medium effect, whereas $g = 0.8$ denotes a large effect.

We assessed heterogeneity, i.e., the extent to which effect sizes vary within our meta-analysis, by studying Chochran's Q [44] and Higgins and Thompson's I^2 [78]. The first one represents the difference between the observed effect sizes and the fixed-effect model estimate of the effect size. When Q is statistically significant, it indicates that the effect sizes are heterogeneous. The latter,

[11] As reported in Section 6.1, a large part of studies included in our corpus followed a within-subject design without a control group, where users were monitored before and after the DSCT was activated. In studies with a control group, e.g., [145], we considered the baseline and the intervention phases of the intervention group, only, to be consistent with the other analyzed papers.

instead, represents the percentage of variability in the effect sizes which is not caused by sampling error, and it is not sensitive to changes in the number of studies in the analyses. Furthermore, it can be interpreted according to specific guidelines. According to Higgins et al. [79], $I^2 = 25\%$ suggests low heterogeneity, $I^2 = 50\%$ suggests moderate heterogeneity, whereas $I^2 = 75\%$ denotes substantial heterogeneity. We also analyzed the confidence intervals of the effect sizes of each included study to spot any extreme effect sizes, i.e., outliers, and we performed an influence analysis by exploiting the Leave-One-Out method [219] and the Graphic Display of Heterogeneity (GOSH) analysis [168]. Our aim was to detect and remove influential cases, i.e., studies that exert a high influence on the overall effect size. Finally, we assessed publication bias by investigating small sample biases [32].

6.2.2 Outliers Detection and Influence Analysis. We began our analysis looking for *outliers* and *influential cases* in our sample. Outliers are studies with extreme effect sizes that may distort the overall effect size. Influential cases, instead, are studies that exert a high influence on the overall effect size. By detecting and eventually removing these studies, we ought to reduce between-study heterogeneity, thus retrieving an estimation of the pooled effect size that is robust enough, i.e., that it does not depend heavily on a small subset of studies.

We defined an outlier as a study whose confidence interval does not overlap with the confidence interval of the pooled effect size. We therefore conducted a first meta-analysis on all the identified studies by inspecting the confidence interval of each effect size, looking for studies reporting extremely small or large effects. Such an analysis identified the study of Foulonneau et al. (S8) [82] as an outlier with an extremely large effect size. The lower bound of its 95% confidence interval ($g = 1.9789$), indeed, was higher than the upper bound of the confidence interval of the pooled effect size ($g = 1.3883$).

After the outlier detection, we conducted a Leave-One-Out analysis [219] to spot influential cases. In this case, we recalculated the same meta-analysis 7 times, each time leaving out one of the 8 studies. Again, we found that the study of Foulonneau et al. (S8) [82] highly influenced the pooled effect size according to different metrics, e.g., the Cook's distance and the covariance ratio. To further explore patterns of heterogeneity in our sample, we also conducted a Graphic Display of Heterogeneity (GOSH) analysis [168], through which the same meta-analysis model is fitted to all possible subsets of the included studies. Such an analysis, in particular, uses supervised machine learning algorithms to detect if the effect sizes in the sample are homogeneous or if there exist specific patterns, e.g., clusters of studies with different effect sizes. The GOSH analysis identified another study with an extreme effect size which might potentially contribute to cluster imbalance, i.e., the study of Park et al. (S6) [171]. The Baujat plot [24] of Figure 10 further confirms the findings of our outliers detection and influence analysis. The plot shows the contribution of each study to the overall heterogeneity (as measured by Cochran's Q) on the horizontal axis, and its influence on the pooled effect size on the vertical axis. We can see that S8 overly contributes to the heterogeneity of the meta-analysis, while S6 has an extreme effect size.

We therefore removed S8 and S6 from the corpus of studies included in our final meta-analysis. The resulting sample ($N = 7$, [105, 112, 114, 145, 159, 214, 223]) was characterized by a low heterogeneity. The Chochran's Q value (4.56) was not statistically significant ($p = 0.6017$), and the Higgins and Thompson's I^2 was 0.0%, with a 95% confidence interval ranging from 0.0% to 61.6%.

6.2.3 Effect of DSCTs Over Time-spent. Table 10 reports the results of the final meta-analysis conducted on the 7 identified studies, while Figure 11 visualize these results by means of a forest plot.

Before analyzing the results, we generated a contour-enhanced funnel plot [175] of the studies included in our meta-analysis in order to assess publication bias (Figure 12). As the chart shows, the studies are partly significant and non-significant, and they lie quite symmetrically around the pooled

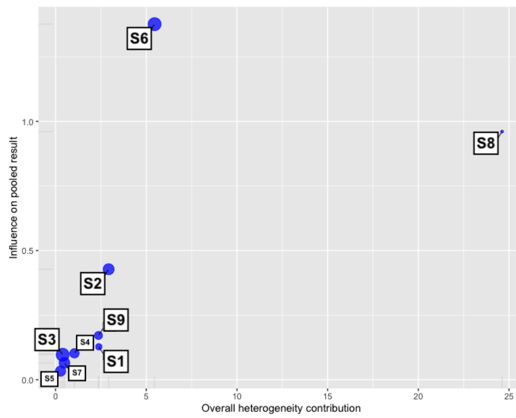

Fig. 10. The Baujat Plot [24] of the studies under analysis, showing the contribution of each study to the overall heterogeneity as measured by Cochran's Q on the horizontal axis, and its influence on the pooled effect size on the vertical axis. The chart suggests that S8 and S6 are outliers.

Table 10. Quantitative results of the performed random-effects-model meta-analysis. The output shows that the estimated effect size of the evaluated DSCTs is $g = 0.47$, and the 95% confidence interval stretches from $g = 0.27$ to $g = 0.68$.

Paper	**Hedges' g**	**Participants [#]**	**95% CI**
Okeke et al. (S5) [114]	0.8325	35	[0.3430;1.3221;]
Whittaker et al. (S3) [223]	0.5863	61	[0.2236;0.9489]
Kim et al. (S7) [105]	0.5340	36	[0.0634; 1.0046]
Ko et al. (S4) [112]	0.4211	27	[-0.1187; 0.9609]
Monge Roffarello et al. (S2) [159]	0.3071	38	[-0.1454;0.7595]
Lyngs et al. (S9) [145]	0.2140	20	[-0.4078; 0.8357]
Tseng et al. (S1) [177]	0.1375	15	[-0.5791;0.8542]
Overall Effect Size	0.4734	255	[0.2657; 0.6811]
Prediction Interval			[0.0332; 0.9136]

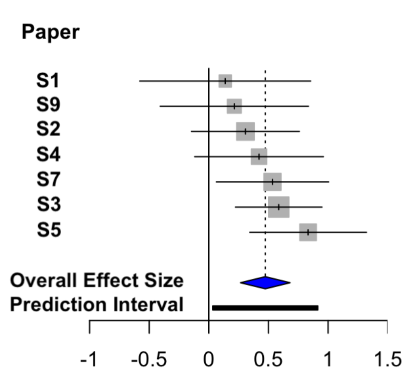

Fig. 11. The forest plot visualizing the effect sizes of the studies included in the meta-analysis, as well as the overall pooled effect size.

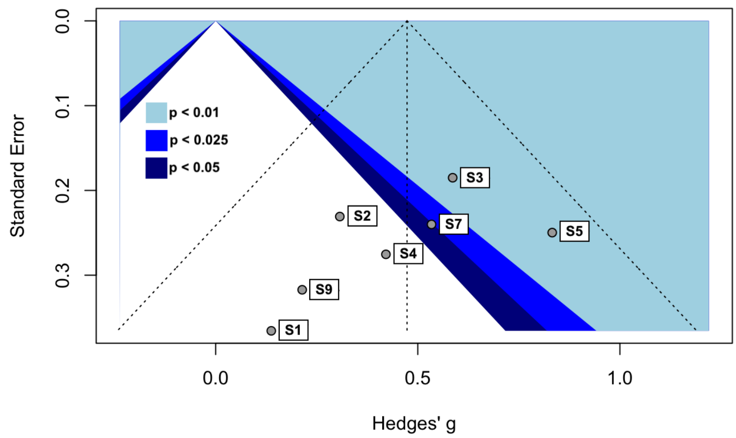

Fig. 12. The contour-enhanced funnel plot [175] of the studies included in our meta-analysis. The studies lie quite symmetrically around the pooled effect size, i.e., the striped line, within the form of the funnel. Furthermore, studies are partly significant (blue backgrounds) and non-significant (white background). We can therefore exclude the the presence of publication bias.

effect size. The overall pooled effect size across all the studies (Table 10), in particular, is $g = 0.4734$, with a 95% confidence interval from 0.2657 to 0.6811 and a total of 255 participants. This means that the evaluated DSCTs had, on average, a small to medium effect on reducing the time spent by users on devices, web sites, and/or mobile applications, according to the Cohen's criteria [45]. As devices such as smartphones and PCs have become an integral part of the daily lives of billions of people, such a relatively small effect of DSCTs may impact a very large population. Furthermore, the cost for reaching such an effectiveness incurs with the design and the implementation of the DSCTs, thereby overcoming the costs associated for delivering individual treatments [1].

Despite the positive aspects, we saw from Section 6.1 that DSCTs are primarily evaluated through short-term studies. The studies involved in our meta-analysis, for instance, had an average duration of 22.57 days ($SD = 10.43$). Therefore, the effect of DSCTs on users' behaviors like time spent on digital devices and services still require a confirmation through longer-term studies, e.g., to understand whether the effectiveness of DSCTs remains constant or varies over time. Furthermore, little is known about whether the new behaviors that users learn by using contemporary DSCTs would survive a break in the use of the tools. A preliminary analysis of the small subset of papers [108, 145, 166] which include a withdrawal phase, i.e., a phase during which interventions are removed, seems rather to suggest that these behaviors actually do not survive without the help of the tool. According to these studies, in particular, only some significant effects, mainly related to self-perception measures like perceived self-control [145], are likely to persist when the DSCT is removed, at least for the limited amount of time of the withdrawal phase. Quantitatively, instead, users gradually return to their old habits without interventions, with the time-spent metric that increases back to nearly the same level of the baseline phase [145, 166]. The reason for this effectiveness degradation can be found in the self-monitoring nature of contemporary DSCTs [159]: unless other strategies like habit-formation [143] are applied, a discontinued use of tools based on self-monitoring strategies like statistics and self-imposed timers makes a changed behavior to slowly return to its pre-intervention level [110].

7 IMPLICATIONS AND FUTURE DIRECTIONS

Our systematic review provide a comprehensive overview of the emerging area of research related to the design of tools for digital self-control and wellbeing. We now discuss the implications of our work, as well as future directions for the digital wellbeing research area, according to three key aspects: (i) overcoming a limited perspective that exclusively focuses on technology overuse and self-monitoring tools, (ii) improving theoretical foundations and evaluations, and (iii) bring ethics in the digital wellbeing discourse and overcome the business model of contemporary tech companies.

7.1 Changing Perspective to Go Beyond Technology Overuse

In this section, we revisit our first three research questions, i.e., underlying motivations (**RQ1**), adopted strategies (**RQ2**), and research challenges (**RQ3**), with the aim of delineating new perspectives and promising research directions in the field of designing technologies for the digital wellbeing. In this context, the negative impact of an excessive technology use on people's digital wellbeing has been largely addressed in previous works. Several experiments and studies from different communities, including HCI, demonstrate that devices and online services like social networks can even promote behaviors that resembles real addictions, to the point of interfering with people's mental health [122] and social interactions [66, 216]. Together with a growing public debate on mainstream media [4, 42], such an attention to the problem of technology overuse has undoubtedly contributed to the emergence of the digital wellbeing topic, and, at the same time, of tools for digital self-control. Our investigation on research goals and targeted digital wellbeing aspects (Section 3), however, demonstrates that the need of designing interventions for digital self-control goes beyond technology overuse. From reducing distractions to avoiding dark patterns, we found a variety of reasons that led researchers to propose and/or design DSCTs. Researchers even motivate their works to reduce side effects related to technology use, e.g., reducing the environmental footprints of the Internet infrastructure. This highlights a very nuanced view of this recent research area, and opens the way to the investigation of different intervention strategies and digital wellbeing aspects.

7.1.1 Designing Better Interventions. While the meta-analysis reported in Section 6.2 shows that implemented DSCTs have a small to medium effect on changing user's behavior, our work also highlights several gaps and challenges characterizing the design of the proposed interventions. Several works included in our analysis, e.g., [19, 159], in particular, relate the ability to exert control over digital devices to the concepts of self-control and self-regulation [144]. The final aim of developing a tool for digital self-control is precisely to assist users to take advantage of (and preferably improve) such an ability. Unfortunately, our analysis on the strategies adopted by contemporary DSCT implementations (*Implemented Tools* corpus, Section 4.1) shows that the way these tools support people's self-regulation is typically based on self-tracking statistics and block/removal strategies, e.g., timers and lockout mechanisms. To be effective, these strategies require the continuous motivation of the user. More importantly, the fact that these interventions are (nearly) always "self-programmed" means that users alone are forced to understand what the causes of their digital wellbeing issues, with the sole help of statistics like screen-time. At the same time, users must also decide what is an appropriate strategy to intervene on their unwanted behaviors, e.g., by selecting a proper time threshold for a usage timer.

We argue that this kind of self-monitoring approaches are not sufficient to effectively improve the impact of digital technologies on people's lives. In this respect, a recent work of Meier and Reinecke [153] suggests the need of going beyond solutions that use screen time as the sole indication of use. The authors suggest that researchers should further differentiate the usage of technology by taking into account the used device, including its features and the type/brand of the used applications, as well as the type of the interaction. In the context of digital wellbeing. The formative study of Monge Roffarello and De Russis [160], for example, demonstrates that users themselves are aware of the fact that distractions can come from any connected device. Using more than one device at the same time, in particular, is a common situation that can result in either positive or negative digital experiences depending on the underlying performed task [160]. Users may indeed have nuanced goals regarding the usage of their devices and services, e.g., to post more content on social networks rather than using them passively [140], or to avoid using smartphones during mealtimes [162]. We therefore call researchers to investigate more "intelligent" DSCTs that are able to analyze and learn from people's multi-device usage patterns, e.g., through machine learning algorithms, with the aim of adapting interventions to various contextual aspects, from the used device to the underlying usage intention [105]. This would enable future DSCTs to provide a *proactive support* to define tailored interventions, by helping those users that do not recognize what are the problems that are negatively influencing their digital wellbeing, e.g., when self-perception diverges from actual use of devices [57, 169]. An example of a research-based tool that has started to pursue such a goal, the only one we found in our analysis on DSCTs strategies (**RQ2**, Section 4.1), is the second version of *Socialize* [188], which is able to detect and notify context-enriched smartphone habits in real-time, with users that can define personalizable just-in-time reminders to avoid such behaviors in the future. Section 5.1 highlights that adaptability could also be useful to reduce attrition. To balance effectiveness and low attrition, in particular, an interesting research direction proposed by Schwartz et al. [198] could be the investigation of *continuously variable* interventions, with DSCTs that are able to progressively increase or decrease the intervention intensity or change "on-the-fly" the adopted strategy by scaling-down strategies that seem to be at high risk of abandonment. Taking inspiration from the DSCTs analyzed in our study, for example, DSCTs could selectively remove parts of the Facebook newsfeed [145], use increasingly complex "interaction restraint" tasks [106, 172], or implement timers before instant blocks [159]. We claim that these solutions would also be useful to promote a step-by-step "learning" processes for their users, a fundamental aspect in the context of behavior change and digital wellbeing [160]: an adaptable DSCT, in particular, could progressively reduce (or increment) its degree of support based on user's achievements, until the user acquires a sufficient level of independence, i.e., it is able to sustain the new behavior without the help of the tool.

7.1.2 From Designing for Avoidance to Re-Designing Interactions. Interestingly, our analysis on strategies also reveals that the interventions that researchers are proposing as future works (Section 4.2) are significantly different from those that have been implemented in the last years, thus highlighting a possible new research direction for the field of interventions for digital self-control. While contemporary DSCTs mainly focuses on blocking and/or mitigating interactions, in particular, several formative studies in our *Qualitative Discussions* corpus explicitly call researchers and developers to explore designs that promote *meaningfulness experiences* and *alternatives*, be they online or offline activities. Some notable exceptions of implemented DSCTs that partially follow such a principle are those that can spur the user to perform an alternative physical or virtual task (see *Socialize* [188] and *Aiki* [88], respectively), and those that aim at re-designing user interfaces and features of existing services, e.g., to reduce dark-patterns in mobile apps [115] or remove "addictive" features like the auto-play of the next video [81].

All in all, focusing on the design of positive and meaningful experiences, e.g., by integrating well-known principles like Slow Design [70] and Mindfulness Design [230], would surely improve the effectiveness of interventions for digital self-control. Unfortunately, re-designing interfaces and features of existing mobile apps or websites is surely more difficult than implementing timers and lockout mechanisms, especially due to technical limitations/restrictions imposed by vendors and providers [80]. While this problem can be circumvented by using alternative approaches (see the community-driven app modification framework proposed by Kollnig et al. [115], that is based on third-parties patches), we argue that more effort should be put into designing DSCTs that can make internal changes to application designs. Furthermore, we envision a future in which digital services promote users' digital wellbeing *by design*, i.e., without the need of an "external" DSCT, with the aim of overcoming the inherent contradiction of designing for digital wellbeing in a business model which currently incentivizes frequent and continuous usage (see Section 7.3.2). This would necessarily require a change in perspective by contemporary tech companies, starting from providing developers with more transparent APIs [80] to finding alternative business models (see Section 7.3.2 for further discussions on this topic).

7.2 Improving Theoretical Foundations and Evaluations

Independently of the chosen perspective, be it designing better interventions or re-designing interactions, the theoretical foundations of the developed solutions and their subsequent evaluations still play a crucial role. In this section, we stem from our findings on these topics (**RQ2**, **RQ3**, **RQ4**) to identify gaps in the analyzed papers and determine how researchers could better integrate theories and effective evaluations in their works.

7.2.1 Mind the Theoretical Gap. Grounding the design of behavior change technologies on well-established behavioral theories is fundamental to generate long-lasting results [178]. Yet, such a theoretical-grounded approach still needs to be further established. On the one hand, some systematic and meta-reviews have identified a general poor use of theory in research investigating digital interventions for behavior change [155, 156], and other reviews have shown that such a trend is common in commercially-available apps, too [48]. On the other hand, the number of HCI researchers that draw on behavioral sciences to inform the design of behavior change technologies is continuously growing [77, 178]. Such a contamination between HCI and behavioral science communities characterizes different domains, from supporting healthy diets [184] to promote physical activity [30]. Yet, the plethora of existing theories makes it difficult for HCI researchers to disentangle the right behavioral strategy to be pursued [178], and research findings mainly remain siloed between the two communities [77]. To mitigate such a problem, the paper "Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research" by Hekler et al. [77] provides HCI researchers with guidance on "interpreting, using, and developing behavioral theories" (p. 3307). In their work, the authors highlight the problems that characterize the use of behavioral theories during the design and the evaluation of behavior change technologies, e.g., gaps between theories and concrete designs and lack of large-scale randomized trials. Furthermore, the authors provide several insightful suggestions to assist HCI researchers to ground their prototypes on behavioral science.

With the theoretical analysis reported in Section 4.3, we analyzed the extent to which drawing on behavioral science is common in the context of DSCTs, and, in particular, whether researchers that are proposing and/or designing tools for digital self-control are taking advantage of the valuable advice from previous studies connecting HCI and behavioral science, e.g., the work of Hecker et al. [77]. Our analysis shows that a large quantity of the analyzed papers (38 out of 62) do not mention any behavioral theory nor construct. Despite not using a theoretical-grounded approach, however, these works often mix together different strategies that take inspiration from a known theory. For instance, while several of the DSCTs included in the *Implemented Tools* corpus allow users to set up conscious behavioral goals, only 3 papers [47, 82, 117] explicitly ground such an intervention in the goal setting theory [135, 136]. Besides goal-advancement features, Section 4.1 shows that a single DSCT often includes several other features, from block/removal to self-tracking. Picking and mixing different strategies without a common theoretical grounding is problematic, since it makes it difficult to understand the effects of particular behavioral theories or techniques [166]. As reported by Hekler et al. [77], moreover, picking only some constructs from a theory makes researchers loose the potential of the full conceptual framework for designing the system, and may lead to methodological flaws in interpreting the validity of the proposed DSCT, e.g., when the tool is inadvertently designed based on constructs that do not work independently but rather relay on other constructs that are not included in the design process.

Despite the advantages of grounding digital interventions on behavioral science, e.g., for achieving higher effectiveness [1], how to best utilize behavioral theories is however challenging. As reported by Hekler et al. [77] and Pindler et al. [178], this is mainly due to the fragmentation that characterizes behavioral science, with an overabundance of different theories, approaches, and techniques that result in redundant constructs differently named depending on the domain of origination [13, 154]. Consequently, many studies in behavior change research that claim to be based on theory do not provide sufficient evidence on how the theory relates to the proposed interventions [156]. Such a confusion also emerges from our DSCTs analysis. In Section 5, indeed, we found that the same intervention, e.g., removing the Facebook newsfeed, is sometimes viewed under different theoretical lenses, e.g., operant conditioning [206] and dual system theory [93, 207], thus making it difficult to understand the real effects of the adopted strategies and understand their long-term impact [178].

All in all, we can conclude that also research on digital self-control tools is subjected to a "theoretical gap [77]" that makes the selection of relevant strategies difficult [164]. To develop better DSCTs and overcome such a gap, we suggest a closer collaboration between HCI researchers and behavioral scientists, as called for by Heckler et al. [77] in the broader context of behavior change research. At the same time, we argue that an effective way to ground DSCTs into behavioral theory is to take advantage of the existing explanatory frameworks that aim to close the described theoretical gap in behavior change research. The Habit Alteration Model (HAM) [178] that has been recently proposed by Pinder et al. [178] is a prominent example. The framework synthesizes dual system theory, goal setting theory, and modern habit theory into an integrative model that simplifies the design of digital behavior change interventions able to form and break habits. Focusing on breaking unwanted digital habits and/or substituting them with alternative (wanted) behaviors, in particular, is a promising approach to support behavior change and ensuring its long-term effects [159]. As reported in our analysis, such an approach has been preliminary investigated in the context of smartphones (see the second version of the *Socialize* tool [188]).

7.2.2 The Need of Long-Term Studies With Control Groups and Diverse Populations. Several studies and experiments have demonstrated that behavior change is a long and complex process. The work of Lally et al. [121], for instance, shows that a new behavior requires from a few weeks to almost a year of repetition to become automatic, with substantial variation at individual level. Similarly, Prochaska and Velicer [182] analyzed previous studies in the context of smoking cessation, showing that an efficient behavior change process may require several years of efforts. Stemming from such findings, Klasnja et al. [111] argued that HCI researchers should not use their short-term studies to make claims about participants changing their habits. Indeed, long-term follow-ups are needed to speculate on more permanent behavior change. More recently, Pinder et al. [178] stated that "short-term studies introducing participants to novel behaviours may provide evidence of a strong intention-behaviour link that does not persist over time" (p. 15:42).

Ideally, the usage of a DSCT gives rise to a behavior change process that should lead people to improve their relationships with technology. Such a "learning" process will become particularly important (and explicit) if researchers will explore habit formation strategies in their DSCTs, as called for by recent reviews in the digital wellbeing context [144, 159]. Unfortunately, the papers included in our *Evaluated Tools* corpus fail to provide sufficient evidence for the long-term effectiveness of the evaluated DSCTs. As Section 6.1 shows, indeed, the median duration of a DSCT study is 21 days, only, with a very limited set of papers that include a follow up interview and/or survey. Consequently, the majority of the small to medium effects reported in the paper under analysis, e.g., those assessing the influence of DSCTs on time-spent (Section 6.2), should be interpreted as short-term results, only. Without accompanying short-term studies with long-term follow ups, in particular, we simply do not know if the described changes in participants' behavior have become persistent [111] or if they are simply related to the fact of using a novel tool, with the reported effects that would therefore not survive a break in the use of the DSCT [145, 166]. As reported in Section 6.2.3, the small subset of studies [108, 145, 166] that monitor participants' behavior during a withdrawal phase, i.e., a phase during which interventions are removed, seem rather to suggest that the effects of contemporary DSCTs are strongly related to an ongoing use of the tool. As such, we argue that, in general, researchers evaluating DSCTs should try to adopt long-term studies to make claims on the real impact of their tools on users' behavior, preferably by using follow-up periods. Even some papers in our corpus, indeed, ackwoledge this urgent challenge (see Section 5). As a reference, researchers might consider the work of Marcus et al. [147] in the context of physical activity, which suggests to evaluate interventions through year(s)-long follow-ups. However, as not all the DSCTs are designed to target the same goals (see Section 3.1), also the proper duration of an evaluation may vary. Indeed, while DSCTs to counter compulsive social network usage should ideally be used continuously, other DSCTs, e.g., those aiming at reducing distractions, may be useful in specific periods of the year, only, e.g., during exams periods for students.

Besides duration, another important factor to be considered when evaluating technology for behavior change is the experimental design [111, 178]. Klasnja et al. [111], for example, state that control groups and very large population samples are necessary to unambiguously demonstrate the *effect* of a technology on behavior change. In this respect, the authors highlight that large Randomized Control Trials (RCT) could allow researchers to demonstrate whether a technology effectively promote people's behavior change. RCTs, i.e., the gold standard of efficacy research in health sciences, have also been recently used for assessing digital behavior change interventions, e.g., in the context of promoting physical activity [85].

Unfortunately, the findings reported in Section 6 demonstrate that there is a general lack of control groups in DSCT studies, with a prevalence of within-subject experiments. While allowing participants to try all the functionality of a DSCT may be useful at early stages of development, e.g., to elicit their qualitative feedback [111], we argue that between-subject experiments with control groups are fundamental to assess the effectiveness of DSCTs and avoid problems like confirmation biases [163]. An example of a study included in our corpus that resembles a Randomized Control Trial is the one reported by Lyngs et al. [145] (see Figure 13 for the adopted procedure). Although the study involves a limited number of participants, especially if compared with traditional clinical RCTs, it compares two randomly-selected groups of participants subjected to different Facebook interventions, i.e., *goal reminders* and *remove newsfeed*, with a group of users that receive a "placebo" intervention, i.e., turning the Facebook background from light grey to white. In line with RCT guidelines [197], the authors also report on several surveys and interviews that have been conducted after each phases of the study, including a 5-month follow up that partially addresses the need for long-term evaluations.

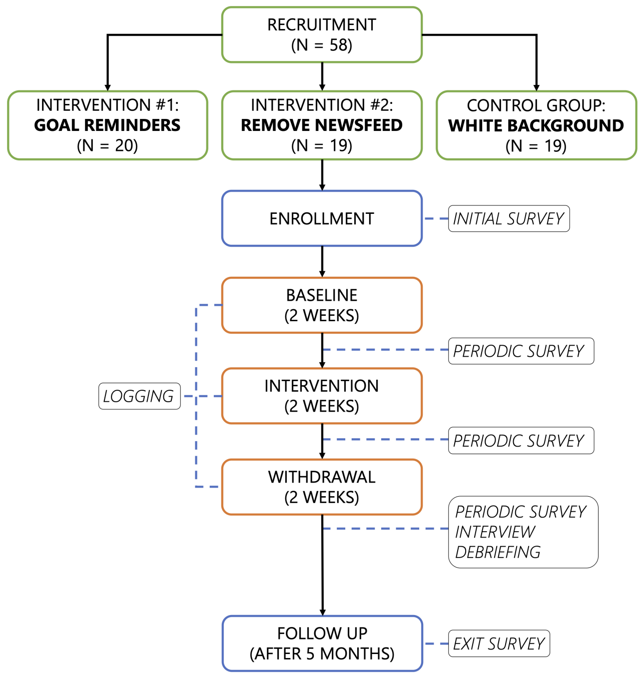

Fig. 13. An example of a study procedure included in our corpus with a control group and a 5-months follow up. The figure is adapted from Lyngs et al. [145].

Finally, the relatively low number of participants that characterize the analyzed studies[12], including the one of Lyngs et al. [145], leads to another important problem characterizing the evaluation of DSCTs. As reported in Section 6.1.1, we found that the recruiting processes described in the papers under analysis suffer from a strong selection bias towards young university students, and, more generally, towards technology savvy users that use devices like PCs and laptops every day, e.g., for studying or working. While considering these categories of users is reasonable, e.g., because they are more exposed to experience firsthand problems like technology overuse, excluding vulnerable categories like teenagers and non-tech savvy users, e.g., middle-aged people that do not use computers as a working tool, creates a gap in the digital wellbeing research context. The former, in particular, need to be involved in the digital wellbeing discourse, since their subjective wellbeing is nowadays undeniably affected, either positively [18] or negatively [220], by the usage of technology. The latter, instead, may experience the same negative consequences of overusing digital devices, e.g., for entertainment purposes, without possessing the necessary skills to proficiently use a DSCT. We therefore call researchers that are designing DSCTs to invest more effort in diversifying their target users and/or recruited participants. We argue that such a comprehensive approach is fundamental to assess whether and how the effects of a DSCT, e.g., those reported in Section 6.2, generalize to wider and diverse populations.

To conclude, we must acknowledge that conducting long-term evaluations of DSCTs by including diverse populations and control groups is not an easy task, e.g., because it requires recruiting several participants that are willing to test (sometimes restrictive) tools for several months. Furthermore, previous work in the HCI area [111, 191] rightly points out that these types of studies are hard

[12] As reported in Section 6.1.1, the median number of participants involved in a DSCT study is 36.

to reconcile with novel and continuously evolving technologies. Apart from requiring a lot of effort, indeed, they often do not allow researchers to understand why an intervention is effective or not [111], therefore undermining the iterations that characterize the early stages of design. These problems apply to DSCTs as well. Digital wellbeing, indeed, is a recent and evolving research topic, and, as reported in our work, researchers in this area are continuously designing and iterating over novel and existing interventions, with a lot of effort that goes into overcoming technical constraints (see Section 5.1).

7.2.3 Standardizing Measures and Rethinking Efficacy. Conducting studies to evaluate the effects of DSCTs on technology use requires first of all to define what the concept of *effectiveness* means in the context of digital wellbeing. As already discussed, the usage of a DSCT should ideally give rise to a behavior change process leading people to improve their relationships with technology. Obviously, such an improvement can be interpreted in many different ways, as technology may influence several aspects of individual's life. In this respect, Section 6.1.2 shows that researchers assess DSCTs by collecting a variety of different measures, ranging from direct assessments of the tools' characteristics to evaluations on how such tools influence users' character traits and feelings, e.g., focus and stress. However, the lack of a standardized definition of effectiveness, as well as the variety of different adopted measures, makes it difficult to compare the effects of different DSCTs. Given the heterogeneous set of measures collected by the study analyzed in this work, for example, the only way to conduct our meta-analysis (Section 6.2) was to focus on the influence of the analyzed DSCTs on the *time spent* by the user on a target device, website, or mobile application. Despite the prevalence of this measure, however, we were only able to include 9 studies in the analysis, e.g., because several studies did not include sufficient statistical data. This requires researchers to provide (well-reported) statistics in their papers, thus allowing other researchers to confirm (or refuse) our findings with larger meta-analysis. To this end, we invite researchers to follow established guidelines and policies, like the ones proposed by the Transparent Statistics in Human–Computer Interaction initiative [8].

That being said, using time spent as a measure for people's digital wellbeing may not be the right choice, and one of the reasons why effects of digital usage on wellbeing are very controversial [123, 124] is because we lack objective measures to gather strong evidence. As already discussed in the previous sections measures like screen time are too coarse, and they do not reflect the variety of goals and different kinds of tech usage of the users. Furthermore, providing users with an indication of their screen time, e.g., for self-regulation purposes, may in turn produce negative reactions [69], thus inducing users to stop using the DSCT [160]. Instead of quantifying their overall screen time, previous work demonstrate that users are more interested in understanding the details of their usage sessions [189]. Unfortunately, our work shows that time spent is still one of the most used metric to evaluate people's digital wellbeing. There is therefore the urgent need for complementary strategies and tools for *measuring* people's digital wellbeing, e.g., through alternative measures, so that researchers could further explore the relationship between wellbeign and technology use. In particular, we argue that HCI researchers should rethink the concept of effectiveness when evaluating their tools for digital self-control. In the broader context of technologies for health behavior change, Klasnja et al. [111] suggest to carry out studies that investigate "how and why a technology works or does not work" (p. 3065). To this end, a possibility is to specifically focus on the constructs that characterize the behavior change strategy implemented by the evaluated tool [111, 191], e.g., by assessing the key mediator variables of a behavioral theory [109]. For those DSCTs like *NUGU* [114] and *FamilyLink* [112] that are informed by the social cognitive theory, for instance, an evaluation revealing an improvement on users' self-efficacy may suggest that the implemented interventions behave as expected. Such an assessment can be performed by measuring the identified construct, e.g., self-efficacy, right *before* and *after* a period of time during which the user adopt the DSCT, e.g., as done by Monge Roffarello and De Russis [159] and Ko et al. [114] through the General Self-Efficacy Scale [92]. Clearly, before rethinking efficacy in the digital wellbeing context and assessing the key mediator variables of the implemented DSCTs, HCI researchers should first focus on closing the theoretical gap discussed in Section 7.2.1.

7.3 Dealing With Ethic and Business Model

Stemming from some interesting strategies proposed in our *Qualitative Discussions* corpus (**RQ2**), and, in particular, taking into account our analysis on challenges and ethical considerations (**RQ3**), this last section discusses other two important aspects that should be further considered in the digital wellbeing context, i.e., the need for ethical guidelines and the role of the HCI community and tech companies.

7.3.1 Bringing Ethics in the Digital Wellbeing Discourse. According to Pinder et al. [178], researchers designing digital interventions for behavior change have a "moral duty" to ensure that their interventions are ethical. Echoing recent reviews in similar domains, e.g., the work of Sanches et al. [191] on HCI and affective health and the work of Thieme et al. [210] on machine learning and mental health, we found a very limited interest in reporting ethical considerations in papers on DSCTs. Section 5.2, in particular, shows that a very limited set of papers explicitly acknowledge having followed ethical guidelines to propose, design, and/or evaluate their interventions for digital self-control. Although not reporting ethical considerations does not mean that ethics was not taken into account, it may however highlight that ethics is sometimes perceived as a secondary aspect. While digital wellbeing cannot be compared, at least in the first instance, with mental health and affective problems, this limited emphasis on ethical considerations is anyway problematic. Indeed, the last few years have seen a growing amount of research attention on the negative aspects of overusing devices like smartphones [129, 217], showing that technology overuse may be associated with negative effects on mental health [122] and social interaction [129]. While debate is ongoing about whether technology overuse should be considered as a real addiction [123], this makes digital wellbeing, and consequently the design of DSCTs, a sensitive topic that may involve vulnerable people. Furthermore, DSCTs collect, by nature, sensitive information like visited websites and contextual information. As reported by Rooksby et al. [189], "data about everyday device use can reveal much about someone's life, to the point that it raises privacy concerns" (p. 293). This opens up new questions related to the *ethics of use* of DSCTs that are still underexplored in the literature, e.g., what happens when workplaces add DSCTs to worker devices or when parents or schools impose DSCTs on children.

All in all, we argue that the first step to bring ethics in the digital wellbeing discourse is therefore to recognize the fact that developing DSCTs countering problematic technology use is, by its nature, a delicate operation that may require researchers to engage with vulnerable users and sensitive information. Recognizing this should encourage designers and researchers to explicitly take into account ethics in the *design* of interventions and, at the same time, to follow ethical research practices when *evaluating* DSCTs. In our opinion, *designing* ethical DSCTs requires researchers to pursue with more effort the core ethical principles of *beneficence* and *non-maleficence*, i.e., to balance benefits against risks and not causing harm. As acknowledged by some papers included in our corpus, indeed, interventions adopted by DSCTs may sometimes have a negative impact on users' wellbeing. The high effectiveness of "punishment" methods, for instance, may in turn provoke high pressure on users [133], since aversive stimuli have the potential of provoking anxiety and defensive aggression [173]. Furthermore, "strict" interventions like locking smartphones may infringe users' autonomy [103]. As advocated by Sanches et al. [191] in the context of HCI and affective health, we already discussed the need of interventions that go beyond "raw" self-monitoring statistics and lockout mechanisms, thus highlighting the need of more innovative DSCTs that help users understand [192] and reflect [193] on their usage patterns to promote positive behavior change (see Section 7.1). Additionally, we see promise in solutions that try to balance risk of abandonment, i.e., attrition, and effectiveness. According to the work of Schwartz et al. [198], "increasing the short-term efficacy of a DSCT is positively correlated with an increasing risk of DSCT abandonment" (p. 2), e.g., because the tool is perceived as too restrictive. This risk-reward tradeoff, that is considered to be a mental heuristic [179], needs to be solved in favor of the user, e.g., by selecting the lowest-risk interventions that still are effective. Furthermore, another important ethical principle that should be considered when designing DSCTs is to take into account individuals' differences (*justice*), e.g., by designing solutions that can adapt to different users' needs, as well as to different users' ages, cultural backgrounds, and educations. DSCTs providing several intervention strategies and complex visualizations, for instance, may be suitable for a user with a technical background, but they might result too complex to be used by average users [16]. Furthermore, the ability to detect the current contextual situations of a user, e.g., to understand if she is currently working or not (see Section 5, *Implementing Accurate Interventions*), may depend on cultural differences in technology adoption [128] and (over)use [225].

As highlighted by our evaluation analysis (Section 6), the need of involving wider and diverse populations is a research practice that also applies to the *evaluation* of DSCTs, e.g., to avoid selection biases. Furthermore, we also argue that researchers should always support users' anonymity, consent, and privacy when evaluating DSCTs. As such tools have the potential to collect sensitive information and to over-restrict users [199], indeed, researchers should use proper informed consent to make explicit how their interventions work before recruiting participants, especially when the implemented strategies are based on System 1 processes[13] [178], e.g., as in [145]. In addition, we believe that researchers should always try to follow (and report) explicit ethical guidelines, e.g., validated by proper Institutional Review Boards, when evaluating DSCTs. In line with previous work on technology to counter affective health problems [191] and self-harm [174], this would require digital wellbeing researchers to take responsibility for recognizing ethical themes *before* carrying out their evaluations, e.g., to screen the recruited participants by taking into account whether and how the implemented DSCTs might provoke excessive harm on some categories of users.

Finally, some discussions included in a small subset of the analyzed papers (see the introductory part of Section 5.2) suggest that ethics in digital wellbeing is not only a responsibility of researchers designing and evaluating DSCTs, but it involves an increasingly important role for *policies* and *regulations*. As DSCTs track their users by collecting sensitive information like device usage, indeed, the risk that these tools are used *against users*, e.g., by violating their privacy, cannot be excluded. As suggested by Widdicks and Pargman [224], this highlights the need of improving and extending existing regulations to impose well-defined limits to the usage of DSCTs, e.g., through additions to the European General Data Protection Regulation (GDPR). Furthermore, we also see the development of new policies in this field as an interesting (and at the same time useful) research direction. How much data are users willing to share with a DSCT in order to improve their digital habits? How much control are users willing to give to a DSCT? Developing policies able to answer these questions would be certainly useful to promote the design of ethical DSCTs that respect users' needs and preferences.

7.3.2 The Role of the HCI Community and Tech Companies. Research works analyzed in this systematic review clearly highlight that an important factor that needs to be considered when

[13] According to the dual system theory [93, 207].

designing interventions for digital self-control is the business model followed by contemporary tech companies. This is an important challenge that needs to be addressed in the near future. Indeed, as reported in Section 5, there is an inherent contradiction between designing interventions to promote digital wellbeing and a business model that aims at maximizing users' attention and engagement to optimize advertising revenue [144], i.e., the so-called Attention Economy [54]. Online services like social networks and video streaming platforms, in particular, are often intentionally designed to maximize frequent and extensive use, to the point of cultivating the idea of promoting automatic and unconscious usage patterns [74, 96], e.g., by exploiting users' psychological vulnerabilities [38].

Unfortunately, our work highlights that the HCI community has so far had an individualistic view of the problem: through DSCTs,researchers investigated how to help users change individually, without taking into account the larger systemic design factors at work. Only recently, researchers [139, 215] have started to wonder what the role of tech companies is in the current "race for digital wellbeing [159]." While tech companies are often blamed for not doing enough against problems like violence and radicalization on social networks, indeed, avoiding problems related to an excessive use of technology has been traditionally considered as a responsibility that belongs solely to the user [215]. As reported in this work, this one-sided view of the digital wellbeing problem is favored by another underlying contradiction that differentiates the digital wellbeing context from other behavior change domains, i.e., the fact that digital devices and services are, at the same time, the source of the problem and the platform with which the interventions are delivered to the user. This is reflected in the variety of tools for digital self-control that have been recently proposed, and on the difficulty of developing cross-platform solutions (see Section 5, *Overcoming Technical Constraints*). Recently, the rising research [123, 194] and main-stream media attention [4, 42] on digital wellbeing has finally spurred tech giants like Google and Apple to acknowledge the problem by introducing self-monitoring usage statistics and tools, as highlighted by the two reviews of contemporary DSCTs included in our corpus (Monge Roffarello and De Russis [159] and Lyings et al. [144]). However, we claim that this only the first step towards a real change in the today's business model: while usage timers and lockout mechanisms may externally apply to many different technological sources, indeed, they do not change the internal attention-capture dark patterns [139] that are often adopted by contemporary websites and mobile applications, e.g., the autoplay of the next video [81, 139].

Just as we believe that the responsibility for digital wellbeing does not lie entirely with users and researchers, we must also acknowledge that a narrative depicting tech companies as the primary source of overuse problems is not useful [215] and mostly speculative [124]. We argue that a real double-sided view of the digital wellbeing topic, one that includes both users, researchers, and tech providers at the same time, is fundamental to promote a positive and meaningful use of technology in the future. Tech companies, for example, could work to minimize the technical constraints that are nowadays hampering the work of HCI researchers and practitioners, e.g., by improving API transparency and by creating permissions to collect usage data [80]. The "Digital Wellbeing Experiments" platform by Google [7] is an example of a promising initiative that goes in this direction: it includes a showcase of guides and open source experiments to assist designers to kick start new ideas and tools for digital self-control. Furthermore, the HCI community and tech providers could work together to find alternative business models that target users' digital wellbeing, rather than users' attention and engagement. A strategy extracted in our review, for example, is to rethink the concept of "relevance" for recommender systems [139], by taking into account when recommendations might promote excessive and compulsive usages of the service. As reported by Lukoff et al. [140] (*Qualitative Discussions* corpus), designing for users' digital wellbeing may initially result in a lower user engagement, thus resulting in a lower business profitability in the short term, but it could also increase user loyalty in the long term. To motivate key stakeholders to support "more responsible and sustainable business [224]", in particular, researchers could engage with policies [224] and standardized definitions: according to Lukoff et al. [139], for example, the design community might encourage tech companies to limit the usage of attention-capture dark patterns by developing "a common language of attention capture dark patterns that recognizes designs that lead to attentional harms" (p.13).

7.4 Limitations

Digital wellbeing is a recent and evolving research area, and new work is constantly emerging. Consequently, we are aware that the corpus of papers included in our systematic literature review cannot be complete. Furthermore, our corpus included papers targeting adults, mainly. Additional research and literature reviews are needed to study similar topics, e.g., technology overuse and related interventions, among children. We also acknowledge that the implications on digital wellbeing reported in this work may be limited by our search methodology, which was exclusively focused on extracting papers proposing or discussing interventions for digital self-control. Finally, the findings reported in Section 6.2 may have been influenced by the fact that the included DSCTs target different categories of users, e.g., students and office workers, and results are constrained to the limited sample size of our meta-analysis and the considered measure, i.e., time-spent. Due to the heterogeneous set of measures included in the papers under analysis and a general tendency towards reporting incomplete statistical data, indeed, we were only able to analyze the effects of 9 DSCTs. Hoping for a standardization of the concept of effectiveness in the digital wellbeing context (see Section 7.2.3), future works would need to confirm (or refuse) our findings on the effects of DSCTs with larger meta-analysis exploiting different (and perhaps more significative) measures.

8 CONCLUSIONS

The growing attention on topics like technology overuse and mobile devices addiction has led to the emergence of a new type of psychological wellbeing related to the impact of the today's information society on people's lives, the so-called *digital wellbeing*. At the same time, HCI researchers and practitioners are developing and evaluating several tools for digital self-control to help users who are struggling with technology use. Through interventions like timers and lockout mechanisms, such tools aim at promoting self-regulation of device and/or specific application use.

In this paper, we have presented the results of the first systematic review of the state-of-the-art literature on digital self-control tools, by showing how researchers are designing, analyzing, and evaluating interventions to promote digital wellbeing. While our work identified a small to medium effect of contemporary digital self-control tools on the time spent by users on technological sources, we also discovered several gaps undermining this promising result. These gaps include a general lack of theory, little considerations of ethical issues and implications, and the prevalent preference towards short-term, within-subject evaluations that unfortunately cannot tell us whether the implemented interventions effectively promote behavior change in the long-term.

As we consider digital self-control tools as a great opportunity to realign technology with users' digital wellbeing, we discussed our findings by highlighting a possible research direction to overcome a limited perspective that exclusively focuses on technology overuse and self-monitoring tools. Finally, we highlighted the possibility of moving towards more ethical, theoretically-grounded tools, and we stressed the need of finding ways to deal with the current business model of contemporary tech-companies, which nowadays incentivizes frequent and continuous usage.

REFERENCES

The papers included in the systematic literature review are highlighted in the References list through a check mark (✓).

[1] 2010. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research* 12, 1 (2010). https://doi.org/10.2196/jmir.1376

[2] 2016. Blocked Facebook ads unblocked, for now. https://www.bbc.com/news/technology-37056013 Accessed: 2021-07-15.

[3] 2016. Latest mobile trends show how people use their devices. https://www.thinkwithgoogle.com/advertising-channels/mobile-marketing/device-use-marketer-tips/ Accessed: 2021-07-12.

[4] 2016. Technology is diminishing us. https://www.theguardian.com/books/2016/dec/03/jonathan-safran-foer-technology-diminishing-us Accessed: 2021-07-15.

[5] 2020. Our commitment to Digital Wellbeing. https://wellbeing.google/ Accessed: 2021-02-12.

[6] 2020. Use Screen Time on your iPhone, iPad, or iPod touch. https://support.apple.com/en-us/HT208982 Accessed: 2021-07-15.

[7] 2021. Digital Wellbeing Experiments. https://experiments.withgoogle.com/collection/digitalwellbeing Accessed: 2021-07-15.

[8] 2022. Transparent Statistics in Human–Computer Interaction. Accessed: 2022-06-16.

[9] Ritu Agarwal and Elena Karahanna. 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly* 24, 4 (2000), 665–694.

[10] Ritu Agarwal, V. Sambamurthy, and Ralph M. Stair. 1997. Cognitive Absorption and the Adoption of New Information Technologies. *Academy of Management Proceedings* 1997, 1 (1997), 293–297. https://doi.org/10.5465/ambpp.1997.4983719

[11] Ryan Ahmed, Alex Chambers, Michael Frontz, and Stephen Voida. 2014. A Tangible Approach to Time Management. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 207–210. https://doi.org/10.1145/2638728.2638794 ✓.

[12] Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 2 (1991), 179 – 211. https://doi.org/10.1016/0749-5978(91)90020-T

[13] Icek Ajzen. 2002. Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior1. *Journal of Applied Social Psychology* 32 (04 2002), 665 – 683. https://doi.org/10.1111/j.1559-1816.2002.tb00236.x

[14] Icek Ajzen and Martin Fishbein. 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin* 84, 5 (1977), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

[15] Morgan G. Ames. 2013. Managing Mobile Multitasking: The Culture of IPhones on Stanford Campus. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1487–1498. https://doi.org/10.1145/2441776.2441945

[16] Ionut Andone, Konrad Blaszkiewicz, Mark Eibes, Boris Trendafilov, Christian Montag, and Alexander Markowetz. 2016. Menthal: Quantifying Smartphone Usage. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 559–564. https://doi.org/10.1145/2968219.2968321 ✓.

[17] Cecilie Schou Andreassen, Ståle Pallesen, and Mark D. Griffiths. 2017. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors* 64 (2017), 287–293. https://doi.org/10.1016/j.addbeh.2016.03.006

[18] Vanessa Apaolaza, Patrick Hartmann, Esteban Medina, Jose M. Barrutia, and Carmen Echebarria. 2013. The relationship between socializing on the Spanish online networking site Tuenti and teenagers' subjective wellbeing: The roles of self-esteem and loneliness. *Computers in Human Behavior* 29, 4 (2013), 1282 – 1289. https://doi.org/10.1016/j.chb.2013.01.002

[19] Julie H. Aranda and Safia Baig. 2018. Toward "JOMO": The Joy of Missing out and the Freedom of Disconnecting. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 19, 8 pages. https://doi.org/10.1145/3229434.3229468 ✓.

[20] Christopher J. Armitage and Mark Conner. 2000. Social cognition models and health behaviour: A structured review. *Psychology & Health* 15, 2 (2000), 173–189. https://doi.org/10.1080/08870440008400299

[21] Albert Bandura. 1986. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

[22] H. L. Barnes and D. H. Olson. 1982. Parent-adolescent communication scale. In *Family inventories: Inventories used in a national survey of families across the family life cycle*, St. Paul: Family Social Science (Ed.). University of Minnesota, 33–48.

[23] Reid Bates. 2004. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning* 27, 3 (2004), 341 – 347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

[24] Bertrand Baujat, Cédric Mahé, Jean-Pierre Pignon, and Catherine Hill. 2002. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in Medicine* 21, 18 (2002). https://doi.org/10.1002/sim.1221

[25] Eric P.S. Baumer, Phil Adams, Vera D. Khovanskaya, Tony C. Liao, Madeline E. Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, Leaving, and (Re)Lapsing: An Exploration of Facebook Non-Use Practices and Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3257–3266. https://doi.org/10.1145/2470654.2466446

[26] Joseph B. Bayer and Scott W. Campbell. 2012. Texting while driving on automatic: Considering the frequency-independent side of habit. *Computers in Human Behavior* 28, 6 (2012), 2083 – 2090. https://doi.org/10.1016/j.chb.2012.06.012

[27] Tom Beauchamp and James Childress. 2019. Principles of Biomedical Ethics: Marking Its Fortieth Anniversary. *The American Journal of Bioethics* 19, 11 (2019), 9–12. https://doi.org/10.1080/15265161.2019.1665402 PMID: 31647760.

[28] A. T. Beck, N. Epstein, G. Brown, and R. A. Steer. 1988. An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology* 56, 6 (1988), 893–897. https://doi.org/10.1037/0022-006X.56.6.893

[29] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. 1961. An inventory for measuring depression. *Archives of General Psychiatry* 4 (1961), 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

[30] Agon Bexheti, Anton Fedosov, Jesper Findahl, Marc Langheinrich, and Evangelos Niforatos. 2015. Re-Live the Moment: Visualizing Run Experiences to Motivate Future Exercises. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 986–993. https://doi.org/10.1145/2786567.2794316

[31] Joël Billieux, Lucien Rochat, Grazia Ceschi, Arnaud Carré, Isabelle Offerlin-Meyer, Anne-Catherine Defeldre, Yasser Khazaal, Chrystel Besche-Richard, and Martial Van der Linden. 2012. Validation of a short French version of the UPPS-P Impulsive Behavior Scale. *Comprehensive psychiatry* 53, 5 (July 2012), 609—615. https://doi.org/10.1016/j.comppsych.2011.09.001

[32] Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. John Wiley & Sons.

[33] Judith Borghouts, Duncan P. Brumby, and Anna L. Cox. 2020. TimeToFocus: Feedback on Interruption Durations Discourages Distractions and Shortens Interruptions. *ACM Transactions on Computer-Human Interaction* 27, 5, Article 32 (Aug. 2020), 31 pages. https://doi.org/10.1145/3396044 ✓.

[34] Harry Brignull. 2010. Dark Patterns. https://darkpatterns.org/ Accessed: 2021-07-12.

[35] Jennings Bryant and Dolf Zillmann. 2002. *Media effects: Advances in theory and research (2nd ed.)*. Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410602428

[36] Moritz Büchi, Noemi Festic, and Michael Latzer. 2019. Digital Overuse and Subjective Well-Being in a Digitized Society. *Social Media + Society* 5, 4 (2019), 2056305119886031. https://doi.org/10.1177/2056305119886031

[37] John R. Buri. 1991. Parental Authority Questionnaire. *Journal of Personality Assessment* 57, 1 (1991), 110–119. https://doi.org/10.1207/s15327752jpa5701_13

[38] Christopher Burr, Nello Cristianini, and James Ladyman. 2018. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines* (2018), 735–774. https://doi.org/10.1007/s11023-018-9479-0

[39] Christopher Burr, Mariarosaria Taddeo, and Luciano Floridi. 2020. The Ethics of Digital Well-Being: A Thematic Review. *Science and Engineering Ethics* (2020), 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

[40] Skinner Burrhus Frederic. 1938. *The behavior of organisms: An experimental analysis.* Appleton-Century.

[41] Marta E. Cecchinato, John Rooksby, Alexis Hiniker, Sean Munson, Kai Lukoff, Luigina Ciolfi, Anja Thieme, and Daniel Harrison. 2019. Designing for Digital Wellbeing: A Research & Practice Agenda. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3298998

[42] Rory Cellan-Jones. 2018. Confessions of a smartphone addict. https://www.bbc.com/news/technology-44972913 Accessed: 2021-07-15.

[43] Alexander Chisler, Yulia Volkova, and Evgeny Pyshkin. 2016. Handle with IT Addiction: A Browser Extension for Overcoming Excessive TV-Series Streaming. In *Proceedings of the 12th Central and Eastern European Software Engineering Conference in Russia* (Moscow, Russia) *(CEE-SECR '16)*. Association for Computing Machinery, New York, NY, USA, Article 2, 5 pages. https://doi.org/10.1145/3022211.3022213 ✓.

[44] William Cochran. 1954. The Combination of Estimates from Different Experiments. *Biometrics* 10, 1 (1954), 101–129. https://doi.org/10.2307/3001666

[45] J. Cohen. 1992. Quantitative methods in psychology: A power primer. *Psychological Bulletin* 112 (1992), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

[46] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A Global Measure of Perceived Stress. *Journal of Health and Social Behavior* 24, 4 (1983), 385–396.

[47] Emily I. M. Collins, Anna L. Cox, Jon Bird, and Cassie Cornish-Tresstail. 2014. Barriers to Engagement with a Personal Informatics Productivity Tool. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (Sydney, New South Wales, Australia) *(OzCHI '14)*. Association for Computing Machinery, New York, NY, USA, 370–379. https://doi.org/10.1145/2686612.2686668 ✓.

[48] Logan T. Cowan, Sarah A. Van Wagenen, Brittany A. Brown, Riley J. Hedin, Yukiko Seino-Stephan, P. Cougar Hall, and Joshua H. West. 2013. Apps of Steel: Are Exercise Apps Providing Consumers With Realistic Expectations?: A Content Analysis of Exercise Apps for Presence of Behavior Change Theory. *Health Education & Behavior* 40, 2 (2013), 133–139. https://doi.org/10.1177/1090198112452126

[49] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[50] M. Csikszentmihalyi. 1975. *Beyond boredom and anxiety - The Experience of Play in Work and Games*. Jossey-Bass.

[51] Mihaly Csikszentmihalyi. 1975. Play and Intrinsic Rewards. *Journal of Humanistic Psychology* 15, 3 (1975), 41–63. https://doi.org/10.1177/002216787501500306

[52] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row.

[53] Melissa A Cyders, Gregory T Smith, Nichea S Spillane, Sarah Fischer, Agnes M Annus, and Claire Peterson. 2007. Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. *Psychological Assessment* 19, 1 (2007), 107–118. https://doi.org/10.1037/1040-3590.19.1.107

[54] Thomas H. Davenport and John C. Beck. 2001. *Attention Economy: Understanding the New Currency of Business*. Harvard Business School Press.

[55] C. De Pasquale, F. Sciacca, and Z. Hichy. 2017. Italian Validation of Smartphone Addiction Scale Short Version for Adolescents and Young Adults (SAS-SV). *Psychology* 8 (2017), 1513–1518. https://doi.org/10.4236/psych.2017.810100

[56] Edward Deci and Richard Ryan. 2012. Motivation, Personality, and Development Within Embedded Social Contexts: An Overview of Self-Determination Theory. *The Oxford Handbook of Human Motivation* (01 2012). https://doi.org/10.1093/oxfordhb/9780195399820.013.0006

[57] David A. Ellis, Brittany I. Davidson, Heather Shaw, and Kristoffer Geyer. 2019. Do smartphone usage scales predict behavior? *International Journal of Human-Computer Studies* 130 (2019), 86 – 92. https://doi.org/10.1016/j.ijhcs.2019.05.004

[58] Greg Feldman, Jeff Greeson, Megan Renna, and Kendra Robbins-Monteith. 2011. Mindfulness predicts less texting while driving among young adults: Examining attention- and emotion-regulation motives as potential mediators. *Personality and Individual Differences* 51, 7 (2011), 856–861. https://doi.org/10.1016/j.paid.2011.07.020

[59] Casey Fiesler. 2020. *Lawful Users: Copyright Circumvention and Legal Constraints on Technology Use*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376745

[60] Michael First, RL Spitzer, ML Gibbon, and Janet Williams. 2002. *Structured clinical interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition.*

[61] B. J. Fogg. 2009. A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, California, USA) *(Persuasive '09)*. Association for Computing Machinery, New York, NY, USA, Article 40, 7 pages. https://doi.org/10.1145/1541948.1541999

[62] Anthony Foulonneau, Gaëlle Calvary, and Eric Villain. 2016. Stop Procrastinating: TILT, Time is Life Time, a Persuasive Application. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Launceston, Tasmania, Australia) *(OzCHI '16)*. Association for Computing Machinery, New York, NY, USA, 508–516. https://doi.org/10.1145/3010915.3010947 ✓.

[63] Freedom. 2020. Freedom to be incredibly productive. https://freedom.to/ Accessed: 2021-01-20.

[64] Hüseyin Ugur Genç and Aykut Coskun. 2020. Designing for Social Interaction in the Age of Excessive Smartphone Use. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376492 ✓.

[65] Margarita Genova, Nermen Ghoniem, and Kevin Doherty. 2021. 'Hung Up': Designing for the Mobile App Engagement University Students Desire. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 339, 7 pages. https://doi.org/10.1145/3411763.3451692 ✓.

[66] Kenneth J. Gergen. 2002. *The challenge of absent presence*. Cambridge University Press, 227–241. https://doi.org/10.1017/CBO9780511489471.018

[67] Jennifer Gerson, Anke C. Plagnol, and Philip J. Corr. 2017. Passive and Active Facebook Use Measure (PAUM): Validation and relationship to the Reinforcement Sensitivity Theory. *Personality and Individual Differences* 117 (2017),

81 – 90. https://doi.org/10.1016/j.paid.2017.05.034
[68] K. Glanz and B. Rimer. 1995. *Theory at a glance: A guide for health promotion practice (second Edition).*
[69] Andrew Griffin. 2018. iOS 12: New iPhone Update's Screen Time Feature Horrifies People as They Discover What They Do on Their Phone. https://www.independent.co.uk/life-style/gadgets-and-tech/news/ios-12-screen-time-iphone-update-latest-ipad-download-apple-a8553146.html Accessed: 2020-07-17.
[70] Barbara Grosse-Hering, Jon Mason, Dzmitry Aliakseyeu, Conny Bakker, and Pieter Desmet. 2013. *Slow Design for Meaningful Interactions.* Association for Computing Machinery, New York, NY, USA, 3431–3440. https://doi.org/10.1145/2470654.2466472
[71] Marco Gui, Marco Fasoli, and Roberto Carradore. 2017. "Digital Well-Being". Developing a New Theoretical Tool For Media Literacy Research. 9 (02 2017). https://doi.org/10.14658/pupj-ijse-2017-1-8
[72] M Harrer, P Cuijpers, TA Furukawa, and DD Ebert. 2019. Doing meta-analysis in R: A hands-on guide. *PROTECT Lab Erlangen* (2019). https://doi.org/10.5281/zenodo.2551803
[73] Bethany Harris, Morgan McCredie, and Sherecce Fields. 2020. Examining the psychometric properties of the Smartphone Addiction Scale and its short version for use with emerging adults in the U.S. *Computers in Human Behavior Reports* 1 (2020), 100011. https://doi.org/10.1016/j.chbr.2020.100011
[74] Tristan Harris. 2017. How Technology Hijacks People's Minds. https://www.huffpost.com/entry/how-technology-hijacks-peoples-minds_b_10155754 Accessed: 2021-02-12.
[75] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909
[76] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9
[77] 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* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3307–3316. https://doi.org/10.1145/2470654.2466452
[78] Julian P. T. Higgins and Simon G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21, 11 (2002), 1539–1558. https://doi.org/10.1002/sim.1186
[79] Julian P. T. Higgins, Simon G. Thompson, Jonathan J. Deeks, and Douglas G. Altman. 2003. Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal* 327, 7414 (2003), 557–560. https://doi.org/10.1136/bmj.327.7414.557
[80] Joshua Hill, Kelly Widdicks, and Mike Hazas. 2020. Mapping the Scope of Software Interventions for Moderate Internet Use on Mobile Devices. In *Proceedings of the 7th International Conference on ICT for Sustainability* (Bristol, United Kingdom) *(ICT4S2020)*. Association for Computing Machinery, New York, NY, USA, 204–212. https://doi.org/10.1145/3401335.3401361 ✓.
[81] Alexis Hiniker, Sharon S. Heung, Sungsoo (Ray) Hong, and Julie A. Kientz. 2018. Coco's Videos: An Empirical Investigation of Video-Player Design Features and Children's Media Use. 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, 1–13. https://doi.org/10.1145/3173574.3173828 ✓.
[82] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4746–4757. https://doi.org/10.1145/2858036.2858403 ✓.
[83] Alexis Hiniker, Kiley Sobel, Hyewon Suh, Yi-Chen Sung, Charlotte P. Lee, and Julie A. Kientz. 2015. Texting While Parenting: How Adults Use Mobile Phones While Caring for Children at the Playground. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 727–736. https://doi.org/10.1145/2702123.2702199 ✓.
[84] Daniel Hintze, Philipp Hintze, Rainhard D. Findling, and René Mayrhofer. 2017. A Large-Scale, Long-Term Analysis of Mobile Device Usage Characteristics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 13 (June 2017), 21 pages. https://doi.org/10.1145/3090078
[85] Robert Hurling, Michael Catt, Marco De Boni, Bruce William Fairley, Tina Hurst, Peter Murray, Alannah Richardson, and Jaspreet Singh Sodhi. 2007. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. *Journal of Medical Internet Research* 9, 2 (2007). https://doi.org/10.2196/jmir.9.2.e7
[86] Edwin Hutchins. 1995. How a cockpit remembers its speeds. *Cognitive Science* 19, 3 (1995), 265 – 288. https://doi.org/10.1016/0364-0213(95)90020-9
[87] NComputing Global Inc. 2020. QualityTime: Discover Your Smartphone Habits. Manage Your Digital Diet. http://www.qualitytimeapp.com/ Accessed: 2020-07-17.

[88] Nanna Inie and Mircea F Lungu. 2021. Aiki - Turning Online Procrastination into Microlearning. 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 369, 13 pages. https://doi.org/10.1145/3411764.3445202 ✓.

[89] Joanna Inthout, John Ioannidis, and George Borm. 2014. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology* 14 (02 2014), 25. https://doi.org/10.1186/1471-2288-14-25

[90] II J. D. Rayburn and Philip Palmgreen. 1984. Merging Uses and Gratifications and Expectancy-Value Theory. *Communication Research* 11, 4 (1984), 537–562. https://doi.org/10.1177/009365084011004005

[91] Se-Hoon Jeong, HyoungJee Kim, Jung-Yoon Yum, and Yoori Hwang. 2016. What type of content are smartphone users addicted to?: SNS vs. games. *Computers in Human Behavior* 54 (2016), 10 – 17. https://doi.org/10.1016/j.chb.2015.07.035

[92] M. Jerusalem and R. Schwarzer. 1992. Self-efficacy as a resource factor in stress appraisal processes. *R. Schwarzer (Ed.), Self-efficacy: Thought control of action* (1992), 195–213.

[93] Daniel Kahneman. 2011. *Thinking, fast and slow.* Farrar, Straus and Giroux.

[94] Daniel Kahneman and Shane Frederick. 2002. *Representativeness Revisited: Attribute Substitution in Intuitive Judgment.* Cambridge University Press, 49–81. https://doi.org/10.1017/CBO9780511808098.004

[95] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5, 1 (March 1991), 193–206. https://doi.org/10.1257/jep.5.1.193

[96] Faltin Karlsen. 2011. Entrapment and Near Miss: A Comparative Analysis of Psycho-Structural Elements in Gambling Games and Massively Multiplayer Online Role-Playing Games. *International Journal of Mental Health and Addiction* (2011), 193–2017. https://doi.org/10.1007/s11469-010-9275-4

[97] Elihu Katz, Jay G. Blumler, and Michael Gurevitch. 1973. Uses and gratifications research. *Public Opinion Quarterly* 37, 4 (1973), 509 – 523. https://doi.org/10.1086/268109

[98] David J. Kavanagh, Jackie Andrade, and Jon May. 2005. Imaginary Relish and Exquisite Torture: The Elaborated Intrusion Theory of Desire. *Psychological Review* 112, 2 (2005), 446–467. https://doi.org/10.1037/0033-295X.112.2.446

[99] Ronald C. Kessler, Lenard Adler, Minnie Ames, Olga Demler, Steve Faraone, Eva Hiripi, Mary J. Howes, Robert Jin, Kristina Secnik, Thomas Spencer, T. Bedirhan Ustun, and Ellen E. Walters. 2005. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine* 35, 2 (2005), 245–256. https://doi.org/10.1017/s0033291704002892

[100] Dongil Kim, Yunhee Lee, Juyoung Lee, JeeEun Karin Nam, and Yeoju Chung. 2014. Development of Korean Smartphone Addiction Proneness Scale for Youth. *PLOS ONE* 9, 5 (05 2014), 1–8. https://doi.org/10.1371/journal.pone.0097920

[101] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 134 (Dec. 2020), 33 pages. https://doi.org/10.1145/3432213 ✓.

[102] Inyeop Kim, Gyuwon Jung, Hayoung Jung, Minsam Ko, and Uichin Lee. 2017. Let's Focus: Location-Based Intervention Tool to Mitigate Phone Use in College Classrooms. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 101–104. https://doi.org/10.1145/3123024.3123165 ✓.

[103] 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, Article 63 (Sept. 2017), 29 pages. https://doi.org/10.1145/3130928 ✓.

[104] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-Device Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 64 (Sept. 2017), 21 pages. https://doi.org/10.1145/3130932 ✓.

[105] 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, Article 16 (March 2019), 29 pages. https://doi.org/10.1145/3314403 ✓.

[106] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300927 ✓.

[107] Sun Mi Kim, Doug Hyun Han, Young Sik Lee, and Perry F. Renshaw. 2012. Combined cognitive behavioral therapy and bupropion for the treatment of problematic on-line game play in adolescents with major depressive disorder. *Computers in Human Behavior* 28, 5 (2012), 1954 – 1959. https://doi.org/10.1016/j.chb.2012.05.015

[108] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York,

NY, USA, 272–283. https://doi.org/10.1145/2858036.2858428 ✓.

[109] Abby King, Daniel Stokols, Emily Talen, Glenn Brassington, and Richard Killingsworth. 2002. Theoretical approaches to the promotion of Physical Activity: Forging a Transdisciplinary paradigm. *American journal of preventive medicine* 23 (09 2002), 15–25. https://doi.org/10.1016/S0749-3797(02)00470-1

[110] 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* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3063–3072. https://doi.org/10.1145/1978942.1979396

[111] 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* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3063–3072. https://doi.org/10.1145/1978942.1979396

[112] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 867–878. https://doi.org/10.1145/2750858.2804283 ✓.

[113] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n' LoL: Group-Based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 998–1010. https://doi.org/10.1145/2858036.2858568 ✓.

[114] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-Based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1235–1245. https://doi.org/10.1145/2675133.2675244 ✓.

[115] Konrad Kollnig, Siddhartha Datta, and Max Van Kleek. 2021. I Want My App That Way: Reclaiming Sovereignty Over Personal Devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 393, 8 pages. https://doi.org/10.1145/3411763.3451632 ✓.

[116] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300560 ✓.

[117] 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, Article 95 (Nov. 2018), 25 pages. https://doi.org/10.1145/3274364 ✓.

[118] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2021. Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently. 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 229, 14 pages. https://doi.org/10.1145/3411764.3445695 ✓.

[119] Min Kwon, Dai-Jin Kim, Hyun Cho, and Soo Yang. 2014. The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents. *PLOS ONE* 8, 12 (12 2014), 1–7. https://doi.org/10.1371/journal.pone.0083558

[120] Min Kwon, Joon-Yeop Lee, Wang-Youn Won, Jae-Woo Park, Jung-Ah Min, Changtae Hahn, Xinyu Gu, Ji-Hye Choi, and Dai-Jin Kim. 2013. Development and Validation of a Smartphone Addiction Scale (SAS). *PLOS ONE* 8, 2 (02 2013), 1–7. https://doi.org/10.1371/journal.pone.0056936

[121] Phillippa Lally, Cornelia H. M. van Jaarsveld, Henry W. W. Potts, and Jane Wardle. 2010. How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology* 40, 6 (2010), 998–1009. https://doi.org/10.1002/ejsp.674

[122] Klodiana Lanaj, Russell E. Johnson, and Christopher M. Barnes. 2014. Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. *Organizational Behavior and Human Decision Processes* 124, 1 (2014), 11 – 23. https://doi.org/10.1016/j.obhdp.2014.01.001

[123] Simone Lanette, Phoebe K. Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is 'Too Much'?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 101 (Nov. 2018), 22 pages. https://doi.org/10.1145/3274370

[124] Simone Lanette and Melissa Mazmanian. 2018. *The Smartphone "Addiction" Narrative is Compelling, but Largely Unfounded.* Association for Computing Machinery, New York, NY, USA, 1–6.

[125] Laura Lascau, Priscilla N. Y. Wong, Duncan P. Brumby, and Anna L. Cox. 2019. Why Are Cross-Device Interactions Important When It Comes To Digital Wellbeing?

[126] Prescott Lecky. 1945. *Self-consistency; a theory of personality.* Island Press.

[127] Heyoung Lee, Heejune Ahn, Samwook Choi, and Wanbok Choi. 2014. The SAMS: Smartphone Addiction Management System and Verification. *Journal of Medical Systems* 38, 1 (Jan. 2014), 1–10. https://doi.org/10.1007/s10916-013-0001-1 ✓.

[128] Sang Gun Lee, Silvana Trimi, and Changsoo Kim. 2013. The impact of cultural differences on technology adoption. *Journal of World Business* 48 (01 2013), 20–29. https://doi.org/10.1016/j.jwb.2012.06.003

[129] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse Among College Students. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2327–2336. https://doi.org/10.1145/2556288.2557366

[130] Alessandro Liberati, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gøtzsche, John P. A. Ioannidis, Mike Clarke, P. J. Devereaux, Jos Kleijnen, and David Moher. 2009. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine* 6, 7 (07 2009), 1–28. https://doi.org/10.1371/journal.pmed.1000100

[131] Sun Lim and Seog Kim. 2018. Validation of a short Korean version of the UPPS-P Impulsive Behavior Scale. *Asia-Pacific Psychiatry* 10 (04 2018), e12318. https://doi.org/10.1111/appy.12318

[132] Jiabao Lin, Shunzhi Lin, Ofir Turel, and Feng Xu. 2020. The buffering effect of flow experience on the relationship between overload and social media users' discontinuance intentions. *Telematics and Informatics* 49 (2020), 101374. https://doi.org/10.1016/j.tele.2020.101374 ✓.

[133] Yikun Liu, Yuan Jia, Wei Pan, and Mark S. Pfaff. 2014. Supporting Task Resumption Using Visual Feedback. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 767–777. https://doi.org/10.1145/2531602.2531710 ✓.

[134] Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: Helping Users with Mobile App Addiction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages. https://doi.org/10.1145/2541831.2541870 ✓.

[135] Edwin Locke. 1990. *A theory of goal setting & task performance.* Prentice-Hall, Inc.

[136] Edwin Locke and Gary Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35Year Odyssey. *American Psychologist - AMER PSYCHOL* 57 (09 2002), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

[137] J. Derek Lomas, Kenneth Koedinger, Nirmal Patel, Sharan Shodhan, Nikhil Poonwala, and Jodi L. Forlizzi. 2017. *Is Difficulty Overrated? The Effects of Choice, Novelty and Suspense on Intrinsic Motivation in Educational Games.* Association for Computing Machinery, New York, NY, USA, 1028–1039.

[138] Olatz Lopez-Fernandez. 2015. Short version of the Smartphone Addiction Scale adapted to Spanish and French: Towards a cross-cultural research in problematic mobile phone use. *Addictive Behaviors* 64 (11 2015). https://doi.org/10.1016/j.addbeh.2015.11.013

[139] 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.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445467 ✓.

[140] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22 (March 2018), 26 pages. https://doi.org/10.1145/3191754 ✓.

[141] A. M. Lund. 2001. Measuring Usability with the USE Questionnaire. *Usability Interfaces* 8, 2 (2001), 3–6.

[142] Aleksandra Luszczynska, Urte Scholz, and Ralf Schwarzer. 2005. The General Self-Efficacy Scale: Multicultural Validation Studies. *The Journal of psychology* 139 (09 2005), 439–57. https://doi.org/10.3200/JRLP.139.5.439-457

[143] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300361

[144] 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.* Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300361 ✓.

[145] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376672 ✓.

[146] T. H. Macan. [n.d.]. Time management: Test of a process model. *Journal of Applied Psychology* 79, 3 ([n. d.]), 381–391. https://doi.org/10.1037/0021-9010.79.3.381

[147] Bess H. Marcus, LeighAnn H. Forsyth, Elaine J. Stone, Patricia M. Dubbert, Thomas L. McKenzie, Andrea L. Dunn, and Steven N. Blair. 2000. Physical activity behavior change: Issues in adoption and maintenance. *Health Psychology* 19, 1 (2000). https://doi.org/10.1037/0278-6133.19.suppl1.32

[148] Claudia Marino, Gianluca Gini, Alessio Vieno, and Marcantonio M. Spada. 2018. A comprehensive meta-analysis on Problematic Facebook Use. *Computers in Human Behavior* 83 (2018), 262 – 277. https://doi.org/10.1016/j.chb.2018.02.009

[149] Gloria Mark, Mary Czerwinski, and Shamsi T. Iqbal. 2018. Effects of Individual Differences in Blocking Workplace Distractions. 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. https://doi.org/10.1145/3173574.3173666 ✓.

[150] Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. 2017. How Blocking Distractions Affects Workplace Focus and Productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 928–934. https://doi.org/10.1145/3123024.3124558 ✓.

[151] Brandon McDaniel and Sarah Coyne. 2014. "Technoference": The Interference of Technology in Couple Relationships and Implications for Women's Personal and Relational Well-Being. *Psychology of Popular Media Culture* 5 (12 2014). https://doi.org/10.1037/ppm0000065

[152] A. Mehrabian and J. Russell. 1974. An approach to environmental psychology.

[153] Adrian Meier and Leonard Reinecke. 0. Computer-Mediated Communication, Social Media, and Mental Health: A Conceptual and Empirical Meta-Review. *Communication Research* 0, 0 (0), 0093650220958224. https://doi.org/10.1177/0093650220958224

[154] Susan Michie, Stefanie Ashford, Falko F. Sniehotta, Stephan U. Dombrowski, Alex Bishop, and David P. French. 2011. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. *Psychology & Health* 26, 1 (2011), 1479–1498. https://doi.org/10.1080/08870446.2010.540664

[155] Susan Michie and Marie Johnston. 2012. Theories and techniques of behaviour change: Developing a cumulative science of behaviour change. *Health Psychology Review* 6, 1 (2012), 1–6. https://doi.org/10.1080/17437199.2012.654964

[156] Susan Michie and Andrew Prestwich. 2010. Are interventions theory-based? Development of a theory coding scheme. *Health Psychology* 29, 1 (2010), 1–8. https://doi.org/10.1037/a0016939

[157] 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, 42 (2011). https://doi.org/10.1186/1748-5908-6-42

[158] David Moher, Alessandroi Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine* 151, 4 (2009), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135 PMID: 19622511.

[159] Alberto Monge Roffarello and Luigi De Russis. 2019. The Race Towards Digital Wellbeing: Issues and Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300616 ✓.

[160] Alberto Monge Roffarello and Luigi De Russis. 2021. *Coping with Digital Wellbeing in a Multi-Device World.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445076 ✓.

[161] David C. Morley, Grayson Lawrence, and Scott Smith. 2016. Virtual Reality User Experience as a Deterrent for Smartphone Use While Driving. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Island, Greece) *(PETRA '16)*. Association for Computing Machinery, New York, NY, USA, Article 67, 3 pages. https://doi.org/10.1145/2910674.2910696 ✓.

[162] Carol Moser, Sarita Y. Schoenebeck, and Katharina Reinecke. 2016. Technology at the Table: Attitudes about Mobile Phone Use at Mealtimes. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1881–1892. https://doi.org/10.1145/2858036.2858357 ✓.

[163] Raymond S. Nickerson. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, 2 (1998), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

[164] Per Nilsen. 2015. Making sense of implementation theories, models and frameworks. *Implementation Science* (2015), 10–53. https://doi.org/10.1186/s13012-015-0242-0

[165] Erick Oduor, Carman Neustaedter, William Odom, Anthony Tang, Niala Moallem, Melanie Tory, and Pourang Irani. 2016. The Frustrations and Benefits of Mobile Device Usage in the Home When Co-Present with Family Members. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. Association for Computing Machinery, New York, NY, USA, 1315–1327. https://doi.org/10.1145/2901790.2901809 ✓.

[166] 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* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3229434.3229463 ✓.

[167] Fabian Okeke, Michael Sobolev, and Deborah Estrin. 2018. Towards A Framework for Mobile Behavior Change Research. In *Proceedings of the Technology, Mind, and Society* (Washington, DC, USA) *(TechMindSociety '18)*. Association for Computing Machinery, New York, NY, USA, Article 27, 6 pages. https://doi.org/10.1145/3183654.3183706 ✓.

[168] Ingram Olkin and Thomas Trikalinos. 2012. GOSH – a graphical display of study heterogeneity. *Research Synthesis Methods* 3 (09 2012). https://doi.org/10.1002/jrsm.1053

[169] Amy Orben and Andrew K. Przybylski. 2019. Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychological Science* 30, 5 (2019), 682–696. https://doi.org/10.1177/0956797619830329

[170] Gábor Orosz, István Tóth-Király, and Beáta Bőthe. 2016. Four facets of Facebook intensity — The development of the Multidimensional Facebook Intensity Scale. *Personality and Individual Differences* 100 (2016), 95 – 104. https://doi.org/10.1016/j.paid.2015.11.038 Dr. Sybil Eysenck Young Researcher Award.

[171] 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.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445489 ✓.

[172] Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188613 ✓.

[173] Sung Yong Park, Sun Mi Kim, Sungwon Roh, Min-Ah Soh, Sang Hoon Lee, Hyungjin Kim, Young Sik Lee, and Doug Hyun Han. 2016. The Effects of a Virtual Reality Treatment Program for Online Gaming Addiction. *Computer Methods and Programs in Biomedicine* 129, C (June 2016), 99–108. https://doi.org/10.1016/j.cmpb.2016.01.015 ✓.

[174] Jessica Pater and Elizabeth Mynatt. 2017. Defining Digital Self-Harm. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1501–1513. https://doi.org/10.1145/2998181.2998224

[175] Jaime L. Peters, Alex J. Sutton, David R. Jones, Keith R. Abrams, and Lesley Rushton. 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology* 61, 10 (2008), 991 – 996. https://doi.org/10.1016/j.jclinepi.2007.11.010

[176] Charlie Pinder, Rowanne Fleck, Rosa Lilia Segundo Díaz, Russell Beale, and Robert J. Hendley. 2016. Accept the Banana: Exploring Incidental Cognitive Bias Modification Techniques on Smartphones. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2923–2931. https://doi.org/10.1145/2851581.2892453

[177] Charlie Pinder, Jose Ignacio Rocca, Benjamin R. Cowan, and Russell Beale. 2019. Push Away the Smartphone: Investigating Methods to Counter Problematic Smartphone Use. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313028 ✓.

[178] Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Transactions on Computer-Human Interaction* 25, 3, Article 15 (June 2018), 66 pages. https://doi.org/10.1145/3196830

[179] Timothy Pleskac and Ralph Hertwig. 2014. Ecologically Rational Choice and the Structure of the Environment. *Journal of experimental psychology. General* 143 (06 2014). https://doi.org/10.1037/xge0000013

[180] G. Porter. 2010. Alleviating the "dark side" of smart phone use. *2010 IEEE International Symposium on Technology and Society* (2010), 435–440.

[181] Chris Preist, Daniel Schien, and Eli Blevis. 2016. *Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure.* Association for Computing Machinery, New York, NY, USA, 1324–1337. https://doi.org/10.1145/2858036.2858378

[182] 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. https://doi.org/10.4278/0890-1171-12.1.38

[183] R Core Team. 2017. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[184] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: Automatic Personalized Health Feedback from User Behaviors and Preferences Using Smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 707–718. https://doi.org/10.1145/2750858.2805840

[185] RescueTime. 2020. Take back control of your time. https://www.rescuetime.com/ Accessed: 2021-02-12.

[186] Felix Riedl, Julia Sageder, and Niels Henze. 2019. Do <i>Knob</i> Disturb: A Tangible Controller for a Distraction-Free Work Environment. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia* (Pisa, Italy) *(MUM '19)*. Association for Computing Machinery, New York, NY, USA, Article 38, 7 pages. https://doi.org/10.1145/3365610.3365638 ✓.

[187] Richard W. Robins, Holly M. Hendin, and Kali H. Trzesniewski. 2001. Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin* 27, 2 (2001), 151–161. https://doi.org/10.1177/0146167201272002

[188] Alberto Monge Roffarello and Luigi De Russis. 2019. Towards Detecting and Mitigating Smartphone Habits. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 149–152. https://doi.org/10.1145/3341162.3343770 ✓.

[189] John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2016. Personal Tracking of Screen Time on Digital Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 284–296. https://doi.org/10.1145/2858036.2858055 ✓.

[190] William Samuelson and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1 (1988), 7–59. https://doi.org/10.1007/BF00055564

[191] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, and Corina Sas. 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300475

[192] Corina Sas and Rohit Chopra. 2015. MeditAid: A Wearable Adaptive Neurofeedback-Based System for Training Mindfulness State. *Personal Ubiquitous Comput.* 19, 7 (Oct. 2015), 1169–1182. https://doi.org/10.1007/s00779-015-0870-z

[193] Corina Sas and Alan Dix. 2009. Designing for Reflection on Experience. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4741–4744. https://doi.org/10.1145/1520340.1520730

[194] Christine Satchell and Paul Dourish. 2009. Beyond the User: Use and Non-use in HCI. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7* (Melbourne, Australia) *(OZCHI '09)*. ACM, New York, NY, USA, 9–16. https://doi.org/10.1145/1738826.1738829

[195] Sarita Yardi Schoenebeck. 2014. Giving up Twitter for Lent: How and Why We Take Breaks from Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 773–782. https://doi.org/10.1145/2556288.2556983 ✓.

[196] Urte Scholz, Benicio Gutiérrez Doña, Shonali Sud, Ralf Schwarzer, and Costa Rica. 2002. Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment* 18, 3 (2002), 242–251. https://doi.org/10.1027/1015-5759.18.3.242

[197] Kenneth F Schulz, Douglas G Altman, and David Moher. 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340 (2010). https://doi.org/10.1136/bmj.c332 arXiv:https://www.bmj.com/content

[198] R.X. Schwartz, Alberto Monge Roffarello, Luigi De Russis, and Panagiotis Apostolellis. 2021. Reducing Risk in Digital Self-Control Tools: Design Patterns and Prototype. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 334, 7 pages. https://doi.org/10.1145/3411763.3451843 ✓.

[199] R. X. Schwartz. 2019. Ulysses' ropes and the inherent limits of digital self-control tools. In *5th International Conference on the History and Philosophy of Computing*.

[200] Ralf Schwarzer, Aristi Born, Saburo Iwawaki, and Young-Min Lee. 1997. The assessment of optimistic self-beliefs: Comparison of the Chinese, Indonesian, Japanese, and Korean versions of the General Self-Efficacy scale. *Psychologia: An International Journal of Psychology in the Orient* 40, 1 (1997), 1–13.

[201] Seekrtech. 2020. Forest - Stay focused, be present. https://www.forestapp.cc/ Accessed: 2020-07-17.

[202] Edwin Shen, Justin Shen, and Tsorng-Lin Chia. 2019. Development of an App to Support Self-Monitoring Smartphone Usage and Healthcare Behaviors in Daily Life. In *Proceedings of the 3rd International Conference on Big Data and*

Internet of Things (Melbourn, VIC, Australia) *(BDIOT 2019)*. Association for Computing Machinery, New York, NY, USA, 29–34. https://doi.org/10.1145/3361758.3361771 ✓.

[203] Muzafer Sherif. 1935. A study of some social factors in perception. *Archives of Psychology (Columbia University)* 27, 187 (1935), 1–60.

[204] 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* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1058–1069. https://doi.org/10.1145/2675133.2675193 ✓.

[205] Katherine W Song, Janaki Vivrekar, Lynn Yeom, Eric Paulos, and Niloufar Salehi. 2021. Crank That Feed: A Physical Intervention for Active Twitter Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 433, 6 pages. https://doi.org/10.1145/3411763.3451817 ✓.

[206] J. E. R. Staddon and D. T. Cerutti. 2003. Operant Conditioning. *Annual Review of Psychology* 54, 1 (2003), 115–144. https://doi.org/10.1146/annurev.psych.54.101601.145124

[207] Fritz Strack and Roland Deutsch. 2004. Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review* 8, 3 (2004), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

[208] A. Tellegen and G. Atkinson. 1974. Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *ournal of Abnormal Psychology* 83, 3 (1974), 268–277. https://doi.org/10.1037/h0036681

[209] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale Univ.Press.

[210] 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 Transactions on Computer-Human Interaction* 27, 5, Article 34 (Aug. 2020), 53 pages. https://doi.org/10.1145/3398069

[211] Virginia Thomas, Margarita Azmitia, and Steve Whittaker. 2016. Unplugged: Exploring the costs and benefits of constant connection. *Computers in Human Behavior* 63 (2016), 540 – 548. https://doi.org/10.1016/j.chb.2016.05.078

[212] Valerie A. Thompson. 2014. Chapter Two - What Intuitions Are… and Are Not. Psychology of Learning and Motivation, Vol. 60. Academic Press, 35 – 75. https://doi.org/10.1016/B978-0-12-800090-8.00002-0

[213] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. *Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300542 ✓.

[214] Vincent W.-S. Tseng, Matthew L. Lee, Laurent Denoue, and Daniel Avrahami. 2019. Overcoming Distractions during Transitions from Break to Work Using a Conversational Website-Blocking System. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300697 ✓.

[215] Ofir Turel and Christopher Ferguson. 2020. Excessive Use of Technology: Can Tech Providers Be the Culprits? *Commun. ACM* 64, 1 (Dec. 2020), 42–44. https://doi.org/10.1145/3392664

[216] Sherry Turkle. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other.* Basic Books, Inc., New York, NY, USA.

[217] Alexander J.A.M. van Deursen, Colin L. Bolle, Sabrina M. Hegner, and Piet A.M. Kommers. 2015. Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in Human Behavior* 45 (2015), 411 – 420. https://doi.org/10.1016/j.chb.2014.12.039

[218] Philippe Verduyn, David Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. Passive Facebook Usage Undermines Affective Well-Being: Experimental and Longitudinal Evidence. *Journal of Experimental Psychology General* (02 2015). https://doi.org/10.1037/xge0000057

[219] Wolfgang Viechtbauer and Mike W.-L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* 1, 2 (2010), 112–125. https://doi.org/10.1002/jrsm.11

[220] Russell M Viner, Aswathikutty Gireesh, Neza Stiglic, Lee D Hudson, Anne-Lise Goddings, Joseph L Ward, and Dasha E Nicholls. 2019. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. *The Lancet Child & Adolescent Health* 3, 10 (2019), 685 – 696. https://doi.org/10.1016/S2352-4642(19)30186-5

[221] Jane Webster and Hayes Ho. 1997. Audience Engagement in Multimedia Presentations. *SIGMIS Database* 28, 2 (April 1997), 63–77. https://doi.org/10.1145/264701.264706

[222] Stephen P. Whiteside and Donald R. Lynam. 2001. The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences* 30, 4 (2001), 669 – 689. https://doi.org/10.1016/S0191-8869(00)00064-7

[223] Steve Whittaker, Vaiva Kalnikaite, Victoria Hollis, and Andrew Guydish. 2016. 'Don't Waste My Time': Use of Time Information Improves Focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1729–1738. https://doi.org/10.1145/2858036.2858193 ✓.

[224] Kelly Widdicks and Daniel Pargman. 2019. Breaking the Cornucopian Paradigm: Towards Moderate Internet Use in Everyday Life. In *Proceedings of the Fifth Workshop on Computing within Limits* (Lappeenranta, Finland) *(LIMITS '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 8 pages. https://doi.org/10.1145/3338103.3338105 ✓.

[225] Zeyang Yang, Kathryn Asbury, and Mark D. Griffiths. 2019. Do Chinese and British University Students Use Smartphones Differently? A Cross-cultural Mixed Methods Study. *International Journal of Mental Health and Addiction* (2019), 644–657. https://doi.org/10.1007/s11469-018-0024-4

[226] Kimberly S. Young. 1998. Internet Addiction: The Emergence of a New Clinical Disorder. *CyberPsychology & Behavior* 1, 3 (1998), 237–244. https://doi.org/10.1089/cpb.1998.1.237

[227] Kimberly S. Young. 2007. Cognitive Behavior Therapy with Internet Addicts: Treatment Outcomes and Implications. *CyberPsychology & Behavior* 10, 5 (2007), 671–679. https://doi.org/10.1089/cpb.2007.9971

[228] Xiabing Zheng and Matthew K.O. Lee. 2016. Excessive use of mobile social networking sites: Negative consequences on individuals. *Computers in Human Behavior* 65 (2016), 65–76. https://doi.org/10.1016/j.chb.2016.08.011

[229] 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. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 3 (April 2021), 30 pages. https://doi.org/10.1145/3449077 ✓.

[230] Bin Zhu, Anders Hedman, and Haibo Li. 2017. *Designing Digital Mindfulness: Presence-In and Presence-With versus Presence-Through.* Association for Computing Machinery, New York, NY, USA, 2685–2695. 10.1145/3025453.3025590