Reducing Risk in Digital Self-Control Tools: Design Patterns and Prototype

R.X. Schwartz
University of Virginia, Charlottesville, Virginia, United States, res7cd@virginia.edu

Alberto Monge Roffarello
Politecnico di Torino, Torino, Italy, alberto.monge@polito.it

Luigi De Russis
Politecnico di Torino, Torino, Italy, luigi.derussis@polito.it

Panagiotis Apostolellis
University of Virginia, Charlottesville, Virginia, United States, panaga@virginia.edu

Many users take advantage of digital self-control tools to self-regulate their device usage through interventions such as timers and lockout mechanisms. One of the major challenges faced by these tools is the user reacting against their self-imposed constraints and abandoning the tool. Although lower-risk interventions would reduce the likelihood of abandonment, previous research on digital self-control tools has left this area of study relatively unexplored. In response, this paper contributes two foundational principles relating risk and effectiveness; four widely applicable novel design patterns for reducing risk of abandonment of digital self-control tools (continuously variable interventions, anti-aging design, obligatory bundling of interventions, and intermediary control systems); and a prototype digital self-control tool that implements these four low-risk design patterns.

CCS CONCEPTS •Human-centered computing~Human computer interaction (HCI)~HCI theory, concepts and models •Human-centered computing~Collaborative and social computing~Collaborative and social computing theory, concepts and paradigms

Additional Keywords and Phrases: Digital self-control tool, DSCT, design pattern, digital wellbeing, online controlled experiment, digital behavior control intervention, DBCI

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 Introduction

In 2009, Google increased annual revenue by $200 million by selecting the highest click-through rate produced among 41 different shades of blue advertising links [1,2]. Similar online controlled experiments, also known as A/B/n tests, split tests, or multivariable tests, are commonplace among digital technology companies [3,4,5]. By testing different interface options among different groups of users, companies may select an interface that optimizes specific user- or business-oriented metrics, such as session length, click-through rate, or sales [6]. Many companies also use personalized recommendation algorithms to select results that optimize metrics of interest [7,8]. These metrics are often improved by relying on previous information from the user, similar users, or the user's social network [9,10]. However, the use of controlled experiments and personalized recommendation algorithms to increase user engagement—in particular, time spent per day on digital platforms—has resulted in a negative response from some health professionals [11,12], news media [13,14], design professionals [15,16], researchers [17,18], and users [19,20,21,22].

At the same time, technical solutions known as digital self-control tools (DSCTs) have allowed millions of users to programmatically manage their engagement with digital technology [23,24,25,26]. DSCTs, as defined in [23], are "self-binding applications that constrain future usage of devices or specific applications." Their features can be grouped into 4 overarching categories of interventions: block/removal, self-tracking, goal advancement, and reward/punishment [24]. Yet the individual nature of DSCTs also poses several challenges. The tools run on the same platforms that they aim to limit, making them subject to inherent external risk [23,27]. DSCTs may over-restrict users in attempts to compensate for a lack of human accountability [23]. Users may abandon DSCTs because the settings of the DSCT do not match their expectations or are too aggressive [28]. Like other behavior control interventions, DSCTs suffer from significant user attrition, ethical concerns, short evaluation periods in published research, and minimal exploration of nonconscious strategies [24,26,28,29,30].

In response to the current state of the DSCT field, this paper outlines a new direction of research that is specifically focused on reducing risk of user abandonment of a given digital self-control tool. We provide a theoretical contribution of two principles that relate risk and effectiveness in DSCTs. We also outline four design patterns for DSCTs that aim to minimize the risk that a given tool is abandoned. Lastly, we describe a prototype of our new digital self-control tool, "Time Sidekick," that incorporates these design patterns.

2 Background and related work

2.1 Designing DSCTs

Previous research on DSCTs and related behavior-change design tools has used dual-process theories to classify and explain interventions [18,24,29,31,32]. In short, dual-process theories are a group of cognitive theories that define two mental processes for behavioral control: System 1 (nonconscious, heuristic, low-effort, and rapid) and System 2 (conscious, analytic, high-effort, and slow) [33,34,35]. Most behavioral reactions begin with System 1 and stem from cognitive biases, habits, or environmental cues [36]. The accessibility of System 1 means that it is often used by default or when individuals lack cognitive capacity or working memory [37]. Instinctual System 1 processes may be influenced in some cases by conscious System 2 goals, yet this influence is limited by the potential for System 1 cues to overwhelm System 2 [36]. System 2 processes may also pass into System 1 control over time [38]. Importantly, targeting System 1 processes for behavior control may reduce reactance to an intervention [32,39].

Along with classifications of System 1/System 2 interventions in DSCTs, previous studies have identified other desired characteristics of DSCTs or similar behavioral change tools. Pinder et al. [26] identifies these desiderata: *low reactance*, *persistence (for habit generation)*, *simplicity, combination of System 1 and 2 control, use of System 1,* *combination of digital and in-person influences, personalization, measurability, voluntariness, disclosure, privacy,* and *consideration of social deception*. Caraban et al. [29] pose behavior-change characteristics of *low reactance, persistence, efficacy*, and *personalization.* Tran et al. [40] note the DSCT goals of *low reactance, persistence*, *subtle influence*, *behavior-goal alignment*, and *meaning*. Schwartz [23] identifies essential limits of DSCTs, which can be translated into desired characteristics of *low reactance*, *alignment with desires of app store owners*, *reliability*, and *voluntariness*. Lanzing [41] recommends *disclosure*. Gulotta et al. [42] recommend *low reactance, low maintenance, proactive design, social features, personalization,* and *periodic goal reflection*. Renfree et al. [43] and Stawarz et al. [30] suggest that behavior change tools should *discourage dependency*. In contrast with the principles of ubiquitous computing and symbiotic systems, which encourage *unobtrusiveness* and *focus on System 1 cognitive processes* [32,44,45,46], Karapanos [47] suggests that behavior change tools should encourage *conscious engagement, social awareness of behavior*, and *habit formation*. Monge Roffarello and De Russis [25] pose the DSCT goals of *personalization, flexibility*, *efficacy*, *reliability*, *social features*, *obtrusiveness,* and *privacy*. Kovacs et al. [28] suggest that DSCTs should be *obtrusive*. In sum, these studies note that DSCTs should meet core principles of minimization of reactance, efficacy, alignment of user behavior with user goals, voluntariness, privacy, and non-deception.

Another group of studies evaluates current DSCT methods. Lyngs et al. [24] review academic and non-academic projects within the DSCT/digital behavior change research area; a review of non-academic project data indicates *few implementations that use nonconscious context-altering strategies*. Monge Roffarello and De Russis [25] review Android DSCTs and note *few DSCTs that automatically assign interventions or redesign the user interface (UI), as well as both positive and negative user reactions to DSCTs*. Their article also tests the DSCT *Socialize* and notes *high abandonment rates of System 2 blocking methods.* Pinder et al. [26] and Caraban et al. [29] review behavior change technologies and both note *high abandonment rates, short study periods,* and *mixed persistence after study periods*. Pinder and Pinder et. al [31,48,49] discuss cognitive bias modification and subliminal priming as potential methods of nonconscious behavior control, yet note *mixed effectiveness* and *ethical concerns*. Barral et al. [50] note *significant ethical concerns regarding "covert" methods*. Kovacs et al. [28] consider the tool *HabitLab* and suggest an *inverse relationship between efficacy and retention in DSCTs*, as well as *user habituation to System 2 interventions*. A more general category of experiments examine the effects of changing or downgrading website elements on various business and user metrics [3,4,51,52,53,54]. These experiments indicate that effects of degradation *may occur, may not occur, may be persistent after degradation is reversed,* or *may require extended time to demonstrate effects.* However, these experiments are not completely applicable to DSCTs because they do not test self-imposed changes.

In all, previous research indicates significant functional, ethical, and pragmatic challenges in the DSCT field. These papers indicate potential opportunities in more effective System 1 (nonconscious) interventions, interventions that are implemented over long periods of time, UI changes, and automatic assignment of interventions. These methods all are encompassed under low-risk interventions, emphasizing long-term reliability over short-term behavioral change. However, low-risk interventions remain underexplored in the DSCT field: the efficacy of System 1 interventions is most reliably shown external to the DSCT field. Also, previous DSCT design strategy has primarily emphasized effectiveness-maximization and has not considered risk-management as a primary goal in DSCT development.

2.2 Understanding risk and reward in DSCTs

How are risk and reward related in the DSCT field? Of particular relevance to this question is the paper by Kovacs et al. [28], who suggest a direct relationship between DSCT effectiveness (the reward, or reduction in daily time on-site) and attrition (risk of abandonment). This correlation is noted in a comparison between static and shuffled System 2 (conscious) behavioral interventions: static System 2 interventions demonstrated lower efficacy and lower attrition, while shuffled System 2 interventions demonstrated higher efficacy and higher attrition. Although the direction of the causal link is unclear, we would propose this as the weak principle of risk and effectiveness in DSCTs: that increasing the short-term efficacy of a DSCT is positively correlated with an increasing risk of DSCT abandonment. This risk-reward tradeoff is noted in other fields and is considered to be a mental heuristic [55].

This relationship may not hold in all cases. In another experiment in the Kovacs et al. study, an explanation was displayed to users alongside shuffled System 1 interventions, which reduced user attrition in comparison to the group where no explanation was provided. Assuming that effectiveness was stable between the two experimental groups, this example would be an example of a Pareto-optimal decrease in risk (reduced risk is gained with no effectiveness lost). This is to be expected as the DSCT field progresses and new risk-reducing measures are created (or improvements in effectiveness are discovered). In predicting the ideal development of the DSCT field, we would propose a strong principle of risk and effectiveness in DSCTs by conjecturing that risk increases with effectiveness given that only the lowest-risk interventions in each effectiveness class are considered. In other words, one should be able to reduce risk in any DSCT by selecting the optimal intervention in a lower class of effectiveness.

Some caveats would apply to this principle as well. As noted in [26,28], it is likely that intervention risk and effectiveness are not wholly generalizable to populations and are affected somewhat by individual characteristics. Based on recent research, it is likely that users would experience different levels of effectiveness based on their usage patterns and relationship to the application or site of interest [56]. Regardless, an estimate of effectiveness and risk could still be modeled for a particular user given results from a previous population. Pairing one intervention with another intervention may also have unpredictable effects on risk and effectiveness: for example, two interventions could act on the same UI element and negate each other or add to each other.

3 Proposed design patterns

Given high rates of attrition from digital self-control tools, minimal exploration of ethical and effective low-risk DSCT strategies, and indications that interventions may not be effective for all user groups, we propose four design patterns to reduce risk of abandonment of digital self-control tools: continuously variable interventions; anti-aging design; obligatory bundling of interventions; and intermediary control systems.

3.1 Design pattern 1: Continuously variable interventions

As mentioned in the Background section, the DSCT field has focused mainly on System 2 interventions, or those that are perceptible to the user. Conscious interventions, such as those that intentionally remove elements, show notifications, or block sites—may be effective, but may also cause negative feelings in users, such as feelings of helplessness or annoyance [28,29]. As such, it would be valuable to minimize effectiveness to avoid risk of abandonment, either before or precisely when users exhibit dissatisfaction or concern.

In response, we propose adopting interventions that are continuously variable, or that can be scaled from a level of 0 (off) to a level of progressively higher effectiveness. For example, a continuously variable intervention could first remove some suggestions instead of the full sidebar of suggestions, show smaller notifications instead of larger notifications, or use delays before blocks. This would allow DSCTs to adopt a risk-minimization principle—to scale intervention levels from a minimal level upwards—in order to reduce risk of abandonment. This technique would also allow flexible scaling-down of interventions that seem to be at high risk of abandonment.

3.2 Design pattern 2: Anti-aging design

DSCTs, like all software, rely on consistent behavior among external resources. This is especially challenging for DSCTs that modify the UIs of websites, operating systems, or other systems (e.g. DSCTs that block elements or change them): they are particularly vulnerable to *software aging*, considered to be the natural degradation of software capabilities, often due to changes in their technical environment [57,58]. DSCTs are also susceptible to *hostile design*, where the systems under modification become intentionally inhospitable (see [59,60]). Lastly, it may be more difficult to develop DSCTs on certain platforms than it is to develop the apps that they aim to restrict [24].

These concerns contribute to the inherent risk taken on by DSCTs as a part of their software development. Although these concerns may not be directly relevant to users, user experience may be indirectly impacted by poor DSCT reliability or by a slow development cycle. As such, the ability to reduce the risk of a varying environment is beneficial for DSCT reliability. This risk-reduction can be considered on at least four different axes: generalizability towards a large user base (e.g. distinct groups that desire varying levels of change to their browsing), generalizability towards a large number of targets (e.g. various websites or various applications), generalizability towards a large number of platforms (e.g. mobile and desktop platforms), and generalizability towards a large number of implementation methods (e.g. using multiple technical methods to control usage). DSCTs can reduce risk on these axes by consolidating their approaches (choosing the simplest method to fulfill an approach) or by adding redundancy to their approaches (choosing multiple equivalent methods to support a single goal). An example of consolidation is using cross-platform libraries instead of platform-specific innovations. An example of redundancy is using a variety of simultaneous methods to locate an element on a page that should be removed, and using a voting process to remove that element (thereby making the DSCT less vulnerable to interface updates) [61].

3.3 Design pattern 3: Obligatory bundling of interventions

Offering multiple interventions (e.g. visualizations, blockers) has been explored in the HabitLab and Socialize tools, as well as in some tools outside of academic studies [24,25,28]. As noted in the Socialize tool, certain subsets of users may choose to reject interventions of a particular type, such as blockers. However, these DSCTs do not require users to adopt certain combinations of interventions. To reduce risk of the failure of one particular intervention, a DSCT could instead obligate users to adopt a bundle of interventions, instead of allowing a user to adopt a single intervention alone. Beyond basic bundling, two “intelligent” strategies for bundling may also be pursued: risk-balancing bundling, or user-aware bundling. Risk-balancing bundling obligates high-risk interventions selected by a user to be offset by lower-risk interventions, because higher-risk interventions are more likely to be abandoned. For example, a user that selects blocking (high-risk) may be forced to also adopt an intervention that makes notifications slightly less attention-grabbing (low-risk). Similarly, low-risk interventions could be bundled with medium- or high-risk interventions, if these higher-risk interventions are more effective. Ultimately, this bundle could improve risk or effectiveness compared to manual selection alone.

Another strategy for bundling, user-aware bundling, selects a minimum number of interventions to be deployed in a bundle, and replaces failing interventions in the bundle in order to shore up overall DSCT functionality. For example, a DSCT where a user that has adopted 3 interventions may find that one of them is not effective (either by indirect method or by direct user feedback). A DSCT that implements user-aware bundling would select this failing intervention for replacement or modification, strengthening the reliability of the tool in achieving the user’s aim and reducing the likelihood of tool failure. This process could be repeated to find more appropriate interventions, ultimately smoothly reducing the risk of abandonment (cf. [28]).

3.4 Design pattern 4: Intermediary control systems

In most digital self-control tools (except for some research studies, e.g. [28]), users have direct control over which types of interventions are implemented [25]. Users often use their intuition to select DSCTs and the interventions within them. At the same time, users are also often disappointed with these results [25,43]. An alternative option would be to prioritize expert selection over user control. Similar to letting a chef select a dish based on a diner’s tastes—or letting a doctor select a treatment based on a patient’s condition—an intermediate control system may only allow certain kinds of interventions to be selected or may simply adopt interventions automatically. These selections may be based on simplified user decisions (such as user input about general goals) or on other data that is not usually available to users, such as behavior logs. Several scenarios are well-suited for intermediate control systems: when users do not know which interventions would be best for their specific needs; when users do not know which interventions are effective out of a large pool of possibilities; when users do not know which interventions would put them at particular risk of tool abandonment; when users hold unreasonable expectations about certain interventions; or when users do not know which interventions conflict.

3.5 Limitations and ethical concerns

Each of these design patterns reduces user autonomy in exchange for a reduction in risk. As mentioned in previous research, it is important to ensure that the user is informed of this exchange—to preserve present autonomy, to allow an adequate decision regarding future dependence on the DSCT, and to help the user develop a valid mental model [26,30,43]. In particular, immediate ethical concerns appear in design patterns 1 and 4, when interventions or changes within them are so minimal so that they cannot be perceived (yet the user may be opposed to the changes or the interventions). To avoid deception, users should agree on clear limits, guides, and prohibitions for app behavior before handing over control to a DSCT, particularly those that use an intermediary control system. In response, the tool should set expectations for behavior and discuss the interventions it uses. It may also be beneficial to allow the user transparency into tool behavior, perhaps by using graphics as well as text.

4 Time Sidekick prototype

4.1 Design and implementation

We developed our prototype of a low-risk DSCT for the Google Chrome browser, “Time Sidekick”, to be evaluated among a preliminary group of users and then in a larger study, open to the public. As recommended by previous researchers, we aim to meet the aforementioned principles of minimization of reactance, efficacy, alignment of user behavior with user goals, voluntariness, privacy, and non-deception. Our tool focuses on interventions that match our four previous design patterns: we implement (1) continuously variable System 1 interventions, such as small delays to webpages or general changes to user interfaces; (2) anti-aging design that emphasizes website-agnostic changes instead of website-specific changes; (3) obligatory bundling of multiple types of interventions, including bundles that slightly delay both dynamically loaded content as well as initially loaded content; and (4) the use of an intermediary control system with a simple user interface. Combined, we anticipate these changes will reduce user involvement in our low-risk DSCT, hopefully also reducing reactance and abandonment of the tool.

Upon installation, the user is presented with a screen that explains the tool. We highlight that the tool is easy to use and how it helps users control which sites they want to use less. Concretely, the initial screen tells the user that the tool makes “small-to-medium changes to [the user’s] browsing experience” including “changes to timing and color, among other tweaks.”

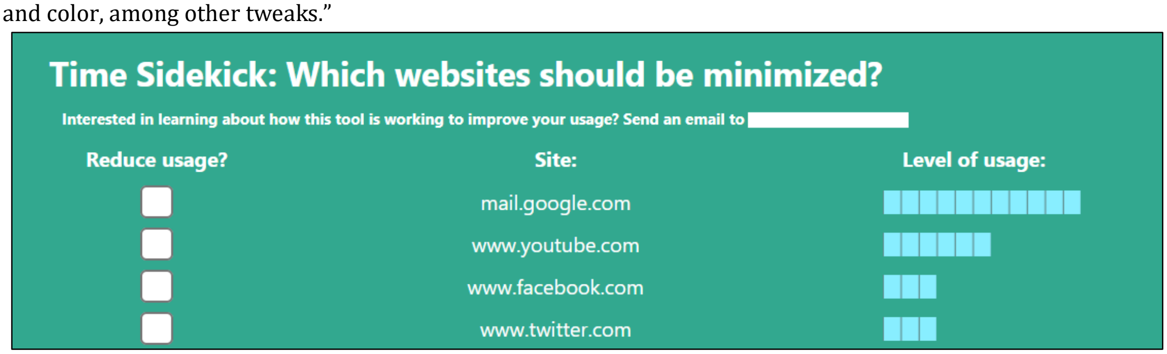

Figure 1. The main user-facing page, showing the intermediary control system interface for Time Sidekick.

Next, the user enters in their personal information, indicating email, age, and gender. We aim to collect this data to determine how usage varies among different user populations, and to contact users if necessary. On this screen, the user is also informed that information about their location and browsing will be collected (this is part of the informed consent process, approved by UVA IRB-SBS #3922). Lastly, the user is presented with the main interface: a list of their most commonly used websites, from the most to least frequently visited, indicated by bars (see Figure 1).

To impose the interventions selected by the control system, the user checks the box at left next to any of the websites they wish to limit (Figure 1). At the top of the page, an email address is also provided through which interested users can also learn about their usage and the interventions imposed. In this way, we aim to provide users transparency into the interventions and the intermediary control system.

Currently, we have implemented two continuously variable interventions in the control system: one intervention that adds a delay before the page loads, and one intervention that adds delays to dynamically loaded content in the page. The influence of delays on user behavior is supported by previous research [51,52]. The two redundant delay interventions we have implemented affect different portions of the browsing experience and are generalizable towards a large number of targets (various websites), as described in Section 3.2. Notably, the initial page load delay is experienced by a user when any single webpage is newly loaded or reloaded, allowing fully website-agnostic behavior. However, because many websites dynamically load content into a single webpage (such as video streaming sites or social media sites), we have added a delay to page requests for dynamic content as well. We predict that the dynamic delays will counteract longer average time spent per page on these sites. By combining both types of delays, we anticipate that Time Sidekick will be more likely to reliably influence usage across a variety of types of websites.

4.2 Proposed evaluation

We record the time and length of the user's visits to any URL, categorized by domain, top-level folder, and a hashed version of the URL. This enables us to investigate the total number of visits and length of time spent per site per day (Figure 2), as well as the median visit length per day, as dependent variables. At the beginning of the study, we will record this usage data over a period of 1-2 weeks in order to establish a baseline level of website usage. During this initial stage, the interventions will not be inactive. In the next stage—the experimental stage—we will separate users into low- and high-effectiveness groups, wherein each user will receive a portfolio of interventions that is either at a low or high level of overall effectiveness. Through our investigation, we aim to understand the impact of the level of effectiveness on our dependent variables (subject to environmental effects), as well as the abandonment of the high-level intervention portfolio versus the low-level intervention portfolio.

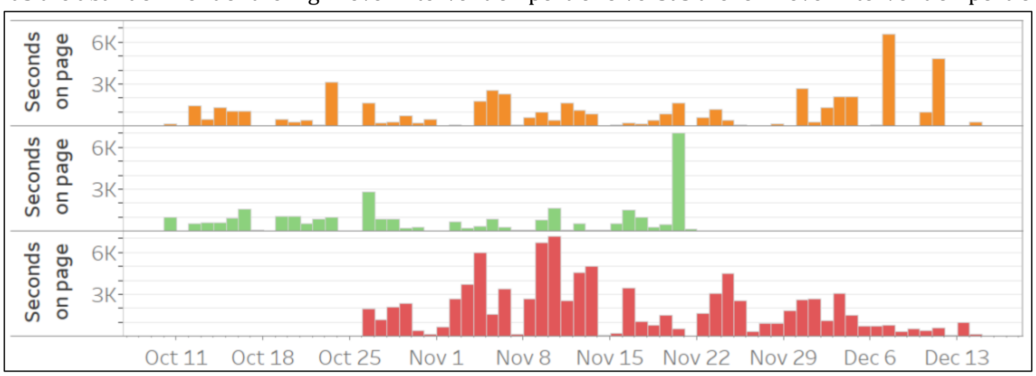

Figure 2: A sample of the researchers' usage data collected by Time Sidekick, showing seconds spent on mail.google.com per day.

In addition to the dependent variables, we also collect other associated user data, including the user's time zone, browser language, and location (city, region, and country). We expect that location will have a moderate effect on the effectiveness of interventions due to local and cultural differences in device and website usage. Beyond baseline differences in usage by location, local or temporal differences in usage may also occur due to COVID-19 outbreaks or lockdowns. Nevertheless, we anticipate that the interventions developed will cause some effect on usage no matter the location.

We also collect information about the number of dynamic delays generated by various types of websites (e.g. video streaming sites, social media sites, email sites) during a browsing session. Preliminary testing has shown that dynamic delays cause unique effects depending on the type of website, and we believe that recording the number of dynamic delays generated by each page browsing session will help us better understand how the dynamic delay intervention works in the wild.

We plan to collect qualitative interview feedback from the participants in our pilot study. In particular, we plan to ask about current usage of websites, target usage of websites, feelings about website usage, perceptions of the sites under modification by the DSCT, and perceptions of the DSCT itself. We aim to conduct exit interviews with any pilot users who abandon Time Sidekick. We hope that this feedback will help us understand which parts of our tool cause reactance in users, as well as which levels of interventions should be avoided in our larger study because they are particularly likely to cause abandonment of the DSCT.

The overall goal of the pilot and large-scale studies is to conduct longer-term analyses of user behavior—on the scale of one month or longer—in order to fill a noted gap in DSCT research [26]. This would help determine if long-term interventions can make behavior changes permanent. We are interested in comparing the attrition rate of users from Time Sidekick with the published attrition rates of other tools.

5 Conclusion and future work

Our DSCT prototype, "Time Sidekick," implements each of the four risk-reducing design patterns we propose in this paper. Time Sidekick departs from previous DSCT research by focusing on low-risk strategies instead of short-term effectiveness. There are some ways in which our design may still be improved: To balance data collection with constraints on data storage and processing power, time spent on a particular website is collected on a second-by-second basis, meaning that views of websites under one second are not recorded. Also, dynamic delays to websites require a minimal initial involvement from researchers to ensure that the delays are not too heavy, because the delay effect is outsized in websites that dynamically load significant amounts of content. Lastly, delays are two of many low-risk continuously variable interventions that we plan to add to the tool, including grayscale filters, changes to website colors, and changes to fonts.

We aim to address research questions focused on low-risk and long-term usage in our initial studies, but there still remain opportunities to explore the topic in more detail. Future research questions, informed by prior work and our own research on understanding the proposed design patterns, could include: In which cases are continuous increases in an intervention level *within* a browsing session as effective as increases in an intervention over many browsing sessions? What is the effect of the rate of increase in the level of a continuously scaled intervention on effectiveness and attrition? As discussed in [28], do low-effectiveness interventions cause low attrition or simply spread attrition over a longer time period? Does decreasing the level of an intervention always reduce attrition? Could increasing the level of an intervention reduce attrition in some cases? We look forward to answering these questions as we continue to develop Time Sidekick.

ACKNOWLEDGEMENTS

The authors thank Paul Humphreys for his advice in the initial stages of this paper. This research is funded by the University of Virginia Human and Machine Intelligence Group and its host, the Humanities Informatics Lab, as well as the University of Virginia Harrison Undergraduate Research Award.

REFERENCES

<bib id="bib1"><number>[1]</number>Alex Hern. 2014. Why Google has 200m reasons to put engineers over designers. The Guardian. Retrieved February 25, 2021 from https://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers</bib>

<bib id="bib2"><number>[2]</number>Laura M. Holson,. 2009. Putting a Bolder Face on Google. The New York Times. Retrieved February 25, 2021 from https://www.nytimes.com/2009/03/01/business/01marissa.html?pagewanted=print</bib>

<bib id="bib3"><number>[3]</number>Alex Deng and Xiaolin Shi. 2016. Data-Driven Metric Development for Online Controlled Experiments: Seven Lessons Learned. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 77–86. https://doi.org/10.1145/2939672.2939700</bib>

<bib id="bib4"><number>[4]</number>Ron Kohavi. 2015. Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 Years. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15, 1–1. https://doi.org/10.1145/2783258.2785464</bib>

<bib id="bib5"><number>[5]</number>Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. 2009. Controlled experiments on the web: survey and practical guide. Data Mining and Knowledge Discovery 18, 1: 140–181. https://doi.org/10.1007/s10618-008-0114-1</bib>

<bib id="bib6"><number>[6]</number>Kerry Rodden, Hilary Hutchinson, and Xin Fu. 2010. Measuring the User Experience on a Large Scale: User-centered Metrics for Web Applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), 2395–2398. https://doi.org/10.1145/1753326.1753687</bib>

<bib id="bib7"><number>[7]</number>Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. Ethics and Information Technology 15, 3: 209–227. https://doi.org/10.1007/s10676-013-9321-6</bib>

<bib id="bib8"><number>[8]</number>Douglas McIlwraith, Haralambos Marmanis, and Dmitry Babenko. 2016. Algorithms of the Intelligent Web. Manning Publications Co., Greenwich, CT, USA.</bib>

<bib id="bib9"><number>[9]</number>Ansgar Koene, Elvira Perez, Christopher James Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, Tom Rodden, and Derek McAuley. 2015. Ethics of Personalized Information Filtering. In Internet Science, Thanassis Tiropanis, Athena Vakali, Laura Sartori and Pete Burnap (eds.). Springer International Publishing, Cham, 123–132. https://doi.org/10.1007/978-3-319-18609-2_10</bib>

<bib id="bib10"><number>[10]</number>Dimitris Paraschakis. 2017. Towards an ethical recommendation framework. In 2017 11th International Conference on Research Challenges in Information Science (RCIS), 211–220. https://doi.org/10.1109/RCIS.2017.7956539</bib>

<bib id="bib11"><number>[11]</number>Royal Society for Public Health. 2017. #StatusOfMind: Social media and young people's mental health and wellbeing. Retrieved February 25, 2021 from https://www.rsph.org.uk/static/uploaded/d125b27c-0b62-41c5-a2c0155a8887cd01.pdf</bib>

<bib id="bib12"><number>[12]</number>Elroy Boers, Mohammad H. Afzali, Nicola Newton, and Patricia Conrod. 2019. Association of Screen Time and Depression in Adolescence. JAMA Pediatrics. https://doi.org/10.1001/jamapediatrics.2019.1759</bib>

<bib id="bib13"><number>[13]</number>Kevin Roose. 2018. Is Tech Too Easy to Use? The New York Times. Retrieved February 25, 2021 from https://www.nytimes.com/2018/12/12/technology/tech-friction-frictionless.html</bib>

<bib id="bib14"><number>[14]</number>Alexis C. Madrigal. 2018. How YouTube's Algorithm Really Works. The Atlantic. Retrieved February 25, 2021 from https://www.theatlantic.com/technology/archive/2018/11/how-youtubes-algorithm-really-works/575212/</bib>

<bib id="bib15"><number>[15]</number>Miguel Helft. 2009. Should Design Be Held Back by a Tyranny of Data? The New York Times. Retrieved February 25, 2021 from https://www.nytimes.com/2009/05/10/business/10ping.html</bib>

<bib id="bib16"><number>[16]</number>Rachel Lerman. 2019. Q&A: Ex-Googler Harris on how tech "downgrades" humans. AP NEWS. Retrieved February 25, 2021 from https://apnews.com/dea7f32d16364c6093f19b938370d600</bib>

<bib id="bib17"><number>[17]</number>Robert Gorwa. 2019. What is platform governance? Information, Communication & Society 22, 6: 854–871. https://doi.org/10.1080/1369118X.2019.1573914</bib>

<bib id="bib18"><number>[18]</number>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 - CHI EA '16, 1389–1397. https://doi.org/10.1145/2851581.2892410</bib>

<bib id="bib19"><number>[19]</number>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 (CHI '13), 3257–3266. https://doi.org/10.1145/2470654.2466446</bib>

<bib id="bib20"><number>[20]</number>Kevin Witzenberger. 2018. The Hyperdodge: How Users Resist Algorithmic Objects in Everyday Life. Media Theory 2, 2: 29–51. Retrieved February 25, 2021 from https://hal.archives-ouvertes.fr/hal-02047585</bib>

<bib id="bib21"><number>[21]</number>Aaron Smith. 2018. Public Attitudes Toward Computer Algorithms. Pew Research Center: Internet, Science & Tech. Retrieved February 25, 2021 from https://www.pewinternet.org/2018/11/16/public-attitudes-toward-computer-algorithms/</bib>

<bib id="bib22"><number>[22]</number>Jesse Fox and Jennifer J. Moreland. 2015. The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. Computers in Human Behavior 45: 168–176. https://doi.org/10.1016/j.chb.2014.11.083</bib>

<bib id="bib23"><number>[23]</number>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. https://doi.org/10.18130/v3-dfzq-ny16</bib>

<bib id="bib24"><number>[24]</number>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. arXiv:1902.00157 [cs]. https://doi.org/10.1145/3290605.3300361</bib>

<bib id="bib25"><number>[25]</number>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 - CHI '19, 1–14. https://doi.org/10.1145/3290605.3300616</bib>

<bib id="bib26"><number>[26]</number>Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. ACM Trans. Comput.-Hum. Interact. 25, 3: 15:1-15:66. https://doi.org/10.1145/3196830</bib>

<bib id="bib27"><number>[27]</number>Jack Nicas. 2019. Apple Cracks Down on Apps That Fight iPhone Addiction. The New York Times. Retrieved February 25, 2021 from https://www.nytimes.com/2019/04/27/technology/apple-screen-time-trackers.html</bib>
<bib id="bib28"><number>[28]</number>Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. Proc. ACM Hum.-Comput. Interact. 2, CSCW: 95:1-95:25. https://doi.org/10.1145/3274364</bib>
<bib id="bib29"><number>[29]</number>Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, 1–15. https://doi.org/10.1145/3290605.3300733</bib>
<bib id="bib30"><number>[30]</number>Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, 2653–2662. https://doi.org/10.1145/2702123.2702230</bib>
<bib id="bib31"><number>[31]</number>Charlie Pinder. 2018. Targeting the automatic: Nonconscious behaviour change using technology. University of Birmingham. Retrieved from https://etheses.bham.ac.uk/id/eprint/8539/1/Pinder18PhD.pdf</bib>
<bib id="bib32"><number>[32]</number>Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. 2015. Mindless computing: designing technologies to subtly influence behavior. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15, 719–730. https://doi.org/10.1145/2750858.2805843</bib>
<bib id="bib33"><number>[33]</number>Jonathan St. B. T. Evans. 2007. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. Annual Review of Psychology 59, 1: 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629</bib>
<bib id="bib34"><number>[34]</number>Shea Houlihan. 2018. Dual-process models of health-related behaviour and cognition: a review of theory. Public Health 156: 52–59. https://doi.org/10.1016/j.puhe.2017.11.002</bib>
<bib id="bib35"><number>[35]</number>Jonathan St B. T. Evans. 2009. How many dual-process theories do we need? One, two, or many? Oxford University Press.</bib>
<bib id="bib36"><number>[36]</number>Daniel Kahneman. 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux.</bib>
<bib id="bib37"><number>[37]</number>Valerie A. Thompson. 2014. Chapter Two - What Intuitions Are… and Are Not. In Psychology of Learning and Motivation, Brian H. Ross (ed.). Academic Press, 35–75. https://doi.org/10.1016/B978-0-12-800090-8.00002-0</bib>
<bib id="bib38"><number>[38]</number>Daniel Kahneman and Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. https://doi.org/10.1017/cbo9780511808098.004</bib>
<bib id="bib39"><number>[39]</number>Jaap Ham and Cees Midden. 2010. Ambient Persuasive Technology Needs Little Cognitive Effort: The Differential Effects of Cognitive Load on Lighting Feedback versus Factual Feedback. In Persuasive Technology (Lecture Notes in Computer Science), 132–142.</bib>
<bib id="bib40"><number>[40]</number>Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, 1–14. https://doi.org/10.1145/3290605.3300542</bib>
<bib id="bib41"><number>[41]</number>Marjolein Lanzing. 2019. "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. Philosophy & Technology 32, 3: 549–568. https://doi.org/10.1007/s13347-018-0316-4</bib>
<bib id="bib42"><number>[42]</number>Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. 2016. Fostering Engagement with Personal Informatics Systems. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16), 286–300. https://doi.org/10.1145/2901790.2901803</bib>
<bib id="bib43"><number>[43]</number>Ian Renfree, Daniel Harrison, Paul Marshall, Katarzyna Stawarz, and Anna Cox. 2016. Don't Kick the Habit: The Role of Dependency in Habit Formation Apps. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16, 2932–2939. https://doi.org/10.1145/2851581.2892495</bib>
<bib id="bib44"><number>[44]</number>Philip Brey. 2005. Freedom and Privacy in Ambient Intelligence. Ethics and Information Technology 7, 3: 157–166. https://doi.org/10.1007/s10676-006-0005-3</bib>
<bib id="bib45"><number>[45]</number>Valeria Orso, Renato Mazza, Luciano Gamberini, Ann Morrison, and Walther Jensen. 2017. Investigating Tactile Stimulation in Symbiotic Systems. In Symbiotic Interaction (Lecture Notes in Computer Science), 137–142.</bib>
<bib id="bib46"><number>[46]</number>Anne-Marie Brouwer, Loïs van de Water, Maarten Hogervorst, Wessel Kraaij, Jan Maarten Schraagen, and Koen Hogenelst. 2018. Monitoring mental state during real life office work. In Symbiotic Interaction: 6th International Workshop, Symbiotic 2017 Eindhoven, The Netherlands, December 18-19, 2017. Revised Selected Papers, 18–29. https://doi.org/10.1007/978-3-319-91593-7_3</bib>
<bib id="bib47"><number>[47]</number>Evangelos Karapanos. 2015. Sustaining User Engagement with Behavior-change Tools. interactions 22, 4: 48–52. https://doi.org/10.1145/2775388</bib>
<bib id="bib48"><number>[48]</number>Charlie Pinder, Jo Vermeulen, Russell Beale, and Robert Hendley. 2015. Exploring Nonconscious Behaviour Change Interventions on Mobile Devices. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15, 1010–1017. https://doi.org/10.1145/2786567.2794319</bib>
<bib id="bib49"><number>[49]</number>Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, Russell Beale, and Robert J. Hendley. 2017. Exploring the feasibility of subliminal priming on smartphones. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17, 1–15. https://doi.org/10.1145/3098279.3098531</bib>
<bib id="bib50"><number>[50]</number>Oswald Barral, Gabor Aranyi, Sid Kouider, Alan Lindsay, Hielke Prins, Imtiaj Ahmed, Giulio Jacucci, Paolo Negri, Luciano Gamberini, David Pizzi, and Marc Cavazza. 2014. Covert Persuasive Technologies: Bringing Subliminal Cues to Human-Computer Interaction. In Persuasive Technology (Lecture Notes in Computer Science), 1–12.</bib>
<bib id="bib51"><number>[51]</number>Miguel Barreda-Ángeles, Ioannis Arapakis, Xiao Bai, B. Barla Cambazoglu, and Alexandre Pereda-Baños. 2015. Unconscious Physiological Effects of Search Latency on Users and Their Click Behaviour. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15, 203–212. https://doi.org/10.1145/2766462.2767719</bib>
<bib id="bib52"><number>[52]</number>Eric Schurman and Jake Brutlag. 2009. The user and business impact of server delays, additional bytes, and http chunking in web search. In Velocity Web Performance and Operations Conference.</bib>
<bib id="bib53"><number>[53]</number>Yang Song, Xiaolin Shi, and Xin Fu. 2013. Evaluating and Predicting User Engagement Change with Degraded Search Relevance. In Proceedings of the 22Nd International Conference on World Wide Web (WWW '13), 1213–1224. https://doi.org/10.1145/2488388.2488494</bib>

<bib id="bib54"><number>[54]</number>Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu. 2012. Trustworthy Online Controlled Experiments: Five Puzzling Outcomes Explained. In KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 9.</bib>
<bib id="bib55"><number>[55]</number>Timothy J. Pleskac and Ralph Hertwig. 2014. Ecologically rational choice and the structure of the environment. Journal of Experimental Psychology: General 143, 5: 2000–2019. https://doi.org/10.1037/xge0000013</bib>
<bib id="bib56"><number>[56]</number>Devansh Saxena, Patrick Skeba, Shion Guha, and Eric P. S. Baumer. 2020. Methods for Generating Typologies of Non/use. Proceedings of the ACM on Human-Computer Interaction 4, CSCW1: 1–26. https://doi.org/10.1145/3392832</bib>
<bib id="bib57"><number>[57]</number>David Lorge Parnas. 1994. Software aging. In Proceedings of the 16th international conference on Software engineering (ICSE '94), 279–287.</bib>
<bib id="bib58"><number>[58]</number>2020. HabitLab. Retrieved from https://chrome.google.com/webstore/detail/habitlab/obghclocpdgcekcognpkblghkedcpdgd?hl=en</bib>
<bib id="bib59"><number>[59]</number>Casey Fiesler. 2020. Lawful Users: Copyright Circumvention and Legal Constraints on Technology Use. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–11. https://doi.org/10.1145/3313831.3376745</bib>
<bib id="bib60"><number>[60]</number>2016. Blocked Facebook ads unblocked, for now. BBC News. Retrieved February 25, 2021 from https://www.bbc.com/news/technology-37056013</bib>
<bib id="bib61"><number>[61]</number>J. von Neumann. 1956. Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components. In Automata Studies. (AM-34), C. E. Shannon and J. McCarthy (eds.). Princeton University Press, Princeton, 43–98. https://doi.org/10.1515/9781400882618-003</bib>